PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/475, 14/435, C12N 15/62, C07K 19/00, C12Q 1/68, C07K 16/18, G01N 33/50, 33/53, A61K 38/18, 48/00

A1

(11) International Publication Number: **WO 99/49041**

(43) International Publication Date: 30 September 1999 (30.09.99)

(21) International Application Number: PCT/US99/06675

(22) International Filing Date: 26 March 1999 (26.03.99)

(30) Priority Data:
60/079,440 26 March 1998 (26.03.98) US

(71) Applicant: THE GOVERNMENT OF THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Office of Technology Transfer, National Institutes of Health, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852-3804 (US).

(72) Inventors: BLAIR, Donald, G.; 8014 Clearfield Road, Frederick, MD 21702-2904 (US). CLAUSEN, Peter, A.; Apartment #3B, 303 Selwyn Drive, Frederick, MD 21701 (US). TOPOL, Lilia, Z.; Apartment #C5, 1411 Key Parkway East, Frederick, MD 21702 (US). MARX, Maria; Institut Curie, F-91405 Orsay (FR). CALOTHY, Georges; Institut Curie, F-91405 Orsay (FR).

(74) Agents: MILLER, Mary, L. et al.; Needle & Rosenberg, P.C., 127 Peachtree Street, N.E., Atlanta, GA 30303 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DRM, A SECRETED PROTEIN WITH CELL GROWTH INHIBITING ACTIVITY, AND RELATED METHODS AND COMPOSITIONS

(57) Abstract

The present invention provides an isolated nucleic acid encoding DRM protein, an isolated DRM polypeptide, and a fusion polypeptide comprising a DRM protein and a green fluorescent protein. The present invention also provides a method of arresting the growth of a cell, comprising administering to the cell an effective amount of DRM protein or an active fragment thereof; a method of inhibiting tumor cell growth, comprising administering to a tumor cell an effective amount of DRM protein or an active fragment thereof; and a method of treating a hyperproliferative cell disorder in a subject diagnosed with a hyperproliferative cell disorder, comprising administering to the subject an effective amount of DRM protein or an active fragment thereof, in a pharmaceutically acceptable carrier. In addition, the present invention provides a method of arresting growth of a cell, comprising administering to the cell an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof; a method of inhibiting tumor cell growth, comprising administering to a tumor cell an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof; and a method of treating a hyperproliferative cell disorder in a subject diagnosed with a hyperproliferative cell disorder, comprising administering to a cell of the subject, in a pharmaceutically acceptable carrier, an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof, under conditions whereby the nucleic acid is expressed in the subject's cell.

06/18/2003, EAST Version: 1.03.0002

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DRM, A SECRETED PROTEIN WITH CELL GROWTH INHIBITING ACTIVITY, AND RELATED METHODS AND COMPOSITIONS

This application claims priority to U.S. provisional application Serial No. 60/0,079,440 filed on March 26, 1998. The 60/079,440 provisional patent application is herein incorporated by this reference in its entirety.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a secreted protein with cell growth inhibiting activity. In particular, the present invention relates to the DRM protein, which is downregulated in transformed cells and which, when overexpressed, can arrest cell growth. The present invention further relates to an enhanced green fluorescent protein (EGFP)/DRM fusion, which imparts stability to the EGFP, thereby enhancing the versatility of EGFP as a research tool.

Background Art

Cell proliferation is determined by a complex and dynamic equilibrium between positive and negative elements signaling the cell to stay in or out of the cycle. The negative elements could be required for an efficient growth shutdown that could end with a reversible (G_0) or irreversible out-of-cycle condition (terminal differentiation, apoptosis, and senescence) (66,67). The exit from the proliferative cell cycle into a reversible quiescence (G_0) is an active process that is not yet well understood at the molecular level. Investigation of G_0-specific gene expression is an important step in studying the mechanism regulating the entrance to quiescence. The nonproliferative state (G_0) in normal cells is characterized by increased expression of a set of genes called *gas* (growth arrest specific) (68). These genes were originally isolated as genes whose expression was increased upon serum starvation or density inhibition (69,70). It has been shown that Gas1, when ectopically expressed, blocks the G_0-to-S phase transition of quiescent fibroblasts (69). The control of cell proliferation occurs mainly in the G1 phase.

Malignant transformation is characterized by alterations in the normal properties of cell growth, adhesion, motility and shape. The multistep nature of this process is now well defined in a number of systems, as well as the fact that genetic changes in specific genes are responsible for both positive and negative contributions to that process. Analysis of the genes involved has identified those which act positively to induce aspects of the transformed state (oncogenes) and more recently, has led to the identification of those which act to block or suppress the malignant phenotype, the so-called tumor-suppressor genes (24). The importance of these genes in maintaining the normal phenotype was first inferred by the fact that in many human tumors their functions have been lost as a consequence of deletion, rearrangement or mutations of both alleles, and indeed the most well-characterized members of this group, represented by Rb, p53, WTI and DCC, were first identified and isolated following pedigree and genetic analyses (34). The frequent physical or functional loss of these tumor-suppressor genes in specific human malignancies was strong evidence that these changes contribute to the development of the neoplastic phenotype.

Loss of function of a particular gene may occur by a variety of mechanisms, including the repression of its expression at the RNA level, and a large number of genes whose expression is repressed either in tumors or in cells transformed by positively acting oncogenes, such as v-*ras*, v-*src* or SV40 T antigen, have been identified. This group includes the retinoic acid receptor (20), α-actinin (13), maspin (44), interferon regulatory factor I (19), tropomyosin (31), as well as the DAN, 322, and rrg genes (8,26,28). Several of these were identified by subtractive hybridization or differential display techniques, which allowed the identification of RNA species whose expression was reduced in transformed cells. In gene transfer experiments, these genes exhibited tumor-suppressive and cell-growth-arrest activities, leading to the hypothesis that the reduced expression or function of certain genes was required for the expression of the transformed phenotype.

The present invention provides a nucleic acid encoding a secreted protein and a secreted protein, designated DRM, with cell growth inhibiting activity and methods for administering the nucleic acid and protein of this invention to arrest cell growth and

treat hyperproliferative cell disorders. The present invention further provides an enhanced green fluorescent protein (EGFP)/DRM fusion which imparts stability to the fluorescence activity of EGFP, thus providing a much more versatile research tool than conventional EGFP.

SUMMARY OF THE INVENTION

The present invention provides an isolated nucleic acid having the nucleotide sequence of SEQ ID NO:2 (human cDNA encoding DRM).The invention also provides an isolated nucleic acid having the nucleotide sequence of SEQ ID NO: 4 (rat cDNA sequence for DRM)

Further provided is an isolated polypeptide having the amino acid sequence of SEQ ID NO:36 (mouse DRM), an isolated nucleic acid encoding the polypeptide and an isolated nucleic acid having the nucleotide sequence of SEQ ID NO:3 (mouse cDNA encoding DRM).

In addition, the present invention provides a method of arresting the growth of a cell, comprising administering to the cell an effective amount of DRM protein or an active fragment thereof; a method of inhibiting tumor cell growth, comprising administering to a tumor cell an effective amount of DRM protein or an active fragment thereof; and a method of treating a hyperproliferative cell disorder in a subject diagnosed with a hyperproliferative cell disorder, comprising administering to the subject an effective amount of DRM protein or an active fragment thereof, in a pharmaceutically acceptable carrier.

In addition, the present invention provides a method of arresting growth of a cell, comprising administering to the cell an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof; a method of inhibiting tumor cell growth, comprising administering to a tumor cell an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof; and a method of treating a hyperproliferative cell disorder in a subject diagnosed with a hyperproliferative cell disorder, comprising administering to a cell of the subject, in a pharmaceutically

acceptable carrier, an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof, under conditions whereby the nucleic acid is expressed in the subject's cell.

Further provided is a method of identifying a subject at risk of developing a hyperproliferative cell disorder, comprising measuring the amount of DRM protein or the amount of nucleic acid encoding DRM in a cell of the subject, whereby an amount of DRM protein or nucleic acid encoding DRM in a cell less than the amount of DRM protein or nucleic acid encoding DRM in a cell of a normal subject identifies a subject at risk of developing a hyperproliferative cell disorder.

The present invention additionally provides a fusion polypeptide comprising a DRM protein and a green fluorescent protein. Also provided is a green fluorescent protein having increased stability, comprising a fusion protein comprising a DRM protein amino acid sequence linked to a green fluorescent protein amino acid sequence.

An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:1 (EGFP/DRM nucleic acid) and a polypeptide having the amino acid of SEQ ID NO:29 (EGFP/DRM amino acid) is also provided.

Further provided is a method of producing a green fluorescent protein having increased stability, comprising the steps of producing a nucleic acid construct whereby a nucleic acid sequence encoding EGFP is positioned upstream and in frame with a nucleic acid encoding DRM or an active fragment thereof; placing the nucleic acid construct into an expression vector; and placing the expression vector into a cell under conditions whereby the nucleic acid of the construct will be expressed, thereby producing a green fluorescent protein having increased stability.

Various other objectives and advantages of the present invention will become apparent from the following detailed description.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

As used herein, "a" or "an" can mean multiples. For example, "a cell" can mean at least one cell.

The present invention is based on the surprising discovery of the secreted protein, DRM, which has been identified to be capable of blocking cell proliferation. The DRM protein, as well as the nucleic acid encoding the DRM protein, can be used in therapeutic applications, to treat hyperproliferative cell disorders, such as cancer. It is further contemplated that the DRM protein and its nucleic acid can be used to identify a subject at risk of developing a hyperproliferative cell disorder, such as cancer.

Thus, the present invention provides an isolated nucleic acid having the nucleotide sequence of SEQ ID NO:2, which encodes the human homologue of the DRM protein having the amino acid sequence of SEQ ID NO:37.

The present invention further provides an isolated polypeptide having the amino acid sequence of SEQ ID NO:36, which is the amino acid sequence of the mouse homologue of DRM. Also provided is an isolated nucleic acid encoding the mouse homologue of DRM and an isolated nucleic acid having the nucleotide sequence of SEQ ID NO:3, which comprises the 5' genomic sequence and the coding sequence of the mouse homologue of DRM. The coding sequence of SEQ ID NO:3 is nucleotides 2201 through 2757. Also provided is a nucleic acid having the nucleotide sequence of SEQ ID NO:4, which encodes the rat homologue of DRM, having the amino acid sequence of SEQ ID NO:38.

"Nucleic acid" as used herein refers to single- or double-stranded molecules which may be DNA, comprised of the nucleotide bases A, T, C and G, or RNA, comprised of the bases A, U (substitutes for T) , C, and G. The nucleic acid may represent a coding strand or its complement. Nucleic acids may be identical in sequence to the sequence which is naturally occurring or may include alternative

codons which encode the same amino acid as that which is found in the naturally occurring sequence (61). Furthermore, nucleic acids may include codons which represent conservative substitutions of amino acids as are well known in the art.

As used herein, the term "isolated" means a nucleic acid separated or substantially free from at least some of the other components of the naturally occurring organism, for example, the cell structural components commonly found associated with nucleic acids in a cellular environment and/or other nucleic acids. The isolation of nucleic acids can therefore be accomplished by techniques such as cell lysis followed by phenol plus chloroform extraction, followed by ethanol precipitation of the nucleic acids (58). The nucleic acids of this invention can be isolated from cells according to methods well known in the art for isolating nucleic acids. Alternatively, the nucleic acids of the present invention can be synthesized according to standard protocols well described in the literature for synthesizing nucleic acids.

The nucleic acid or fragment thereof of this invention can be used as a probe or primer to identify the presence of a nucleic acid encoding the DRM polypeptide in a sample. Thus, the present invention also provides a nucleic acid, which can be the entire complementary sequence to the nucleic acid coding sequence of the DRM protein or a fragment thereof comprising at least eight contiguous nucleotides having sufficient complementarity to the DRM-encoding nucleic acid of this invention to selectively hybridize with the DRM-encoding nucleic acid of this invention under stringent conditions as described herein and which does not hybridize with nucleic acids which do not encode DRM, under stringent conditions.

"Stringent conditions" refers to the hybridization conditions used in a hybridization protocol or in the primer/template hybridization in a polymerase chain reaction (PCR) protocol. In general, these conditions should be a combination of temperature and salt concentration for hybridizing and washing chosen so that the denaturation temperature is approximately 5-20°C below the calculated T_m (melting/denaturation temperature) of the hybrid under study. The temperature and salt conditions are readily determined empirically in routine, preliminary experiments in

which samples of reference nucleic acid are hybridized to the primer nucleic acid of interest and then amplified under conditions of different stringencies. The stringency conditions are readily tested and the parameters altered are readily apparent to one skilled in the art. For example, $MgCl_2$ concentrations used in PCR buffer can be altered to increase the specificity with which the primer binds to the template, but the concentration range of this compound used in hybridization reactions is narrow and therefore, the proper stringency level is easily determined. For example, hybridizations with oligonucleotide probes which are 18 nucleotides in length can be done at 5-10°C below the estimated T_m in 6X SSPE, then washed at the same temperature in 2X SSPE (62). The T_m of such an oligonucleotide can be estimated by allowing 2°C for each A or T nucleotide and 4°C for each G or C. An 18 nucleotide probe of 50% G+C would, therefore, have an approximate T_m of 54°C. Likewise, the starting salt concentration of an 18 nucleotide primer or probe would be about 100-200 mM. Thus, stringent conditions for such an 18 nucleotide primer or probe would be a T_m of about 54°C and a starting salt concentration of about 150 mM and would be modified accordingly by routine, preliminary experiments. T_m values can also be calculated for a variety of conditions utilizing commercially available computer software (e.g., OLIGO®).

Modifications to the nucleic acids of the invention are also contemplated, provided that the essential structure and function of the polypeptide encoded by the nucleic acids is maintained. Likewise, fragments used as primers can have substitutions, provided that a sufficient number of complementary bases exist to allow for selective amplification, as would be determined by routine experimentation (64). In addition, nucleic acid fragments used as probes can have substitutions, provided that enough complementary bases exist to allow for hybridization with the reference sequence to be distinguished from hybridization with other sequences, as would be determined by routine experimentation.

The nucleic acids of this invention can be used as probes, for example, to screen genomic or cDNA libraries or to identify complementary sequences by Northern and Southern blotting. The nucleic acids of this invention can also be used a primers, for

example, to transcribe cDNA from RNA and to amplify DNA according to standard amplification protocols, such as PCR, which are well known in the art.

Thus, the present invention further provides a method of detecting and/or quantitating the expression of a nucleic acid encoding the DRM protein in cells in a biological sample by detecting and/or quantitating DNA and/or mRNA which encodes the DRM protein in the cells comprising the steps of: contacting the cells with a detectably labeled nucleic acid probe that hybridizes, under stringent conditions, with DNA and/or mRNA encoding the DRM protein and detecting and/or quantitating the DNA and/or mRNA hybridized with the probe. The mRNA of the cells in the biological sample can be contacted with the probe and detected and/or quantitated according to protocols standard in the art for detecting and quantitating mRNA, including, but not limited to, Northern blotting, dot blotting, ELISPOT assay and PCR amplification. The DNA of the cells in the biological sample can contacted with the probe and detected and/or quantitated according to protocols standard in the art for detecting and quantitating DNA, including, but not limited to, Southern blotting, dot blotting, ELISPOT assay and PCR amplification. The detection and/or quantitation of DNA or mRNA encoding DRM can be used to identify cells which are undergoing, or about to undergo hyperproliferation (i.e., cells which are cancerous or pre-cancerous), as described further below.

The nucleic acid encoding the polypeptide DRM of this invention can be part of a recombinant nucleic acid comprising any combination of restriction sites and/or functional elements as are well known in the art which facilitate molecular cloning and other recombinant DNA manipulations. Thus, the present invention further provides a recombinant nucleic acid comprising the nucleic acid encoding the DRM protein of the present invention. In particular, the isolated nucleic acid encoding DRM and/or a recombinant nucleic acid comprising a nucleic acid encoding DRM can be present in a vector and the vector can be present in a cell, which can be an *in vivo* cell, an *ex vivo* cell, a cell cultured *in vitro* or a cell in a transgenic non-human animal.

Thus, the present invention further provides a vector comprising a nucleic acid encoding DRM. The composition can be in a pharmaceutically acceptable carrier. The vector can be an expression vector which contains all of the genetic components required for expression of the nucleic acid encoding DRM in cells into which the vector has been introduced, as are well known in the art. The expression vector can be a commercial expression vector or it can be constructed in the laboratory according to standard molecular biology protocols. The expression vector can comprise viral nucleic acid including, but not limited to, adenovirus, retrovirus and/or adeno-associated virus nucleic acid. The nucleic acid or vector of this invention can also be in a liposome or a delivery vehicle which can be taken up by a cell via receptor-mediated or other type of endocytosis.

The present invention further provides a method of producing the polypeptide DRM, comprising culturing the cells of the present invention which contain a nucleic acid encoding the polypeptide DRM under conditions whereby the polypeptide DRM is produced. Conditions whereby the polypeptide DRM is produced can include the standard conditions of any expression system, either *in vitro* or *in vivo*, in which the polypeptides of this invention are produced in functional form. For example, protocols describing the conditions whereby nucleic acids encoding the DRM proteins of this invention are expressed are provided in the Examples section herein. The polypeptide DRM can be isolated and purified from the cells according to methods standard in the art.

With regard to the polypeptides of this invention, as used herein, "isolated" and/or "purified" means a polypeptide which is substantially free from the naturally occurring materials with which the polypeptide is normally associated in nature. Also as used herein, "polypeptide" refers to a molecule comprised of amino acids which correspond to those encoded by a nucleic acid. The polypeptides of this invention can consist of the entire amino acid sequence of the DRM protein or fragments thereof. The polypeptides or fragments thereof of the present invention can be obtained by isolation and purification of the polypeptides from cells where they are produced naturally or by expression of exogenous nucleic acid encoding the DRM polypeptide.

Fragments of the DRM polypeptide can be obtained by chemical synthesis of peptides, by proteolytic cleavage of the polypeptide and by synthesis from nucleic acid encoding the portion of interest. For example, fragments of the DRM polypeptide can comprise the amino acid sequence encoded by nucleotides 4689 through 5147 of SEQ ID NO:5; nucleotides 1339 through 1815 of SEQ ID NO:6; nucleotides 4683 through 5129 of SEQ ID NO:7; nucleotides 4683 through 5033 of SEQ ID NO:8; and nucleotides 4683-5033 of SEQ ID NO:9. The polypeptide may include conservative substitutions where a naturally occurring amino acid is replaced by one having similar properties. Such conservative substitutions do not alter the function of the polypeptide (63).

Thus, it is understood that, where desired, modifications and changes may be made in the nucleic acid and/or amino acid sequence of the DRM polypeptides of the present invention and still obtain a protein having like or otherwise desirable characteristics. Such changes may occur in natural isolates or may be synthetically introduced using site-specific mutagenesis, the procedures for which, such as mis-match polymerase chain reaction (PCR), are well known in the art.

For example, certain amino acids may be substituted for other amino acids in a DRM polypeptide without appreciable loss of functional activity. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a DRM amino acid sequence (or, of course, the underlying nucleic acid sequence) and nevertheless obtain a DRM polypeptide with like properties. It is thus contemplated that various changes may be made in the amino acid sequence of the DRM polypeptide (or underlying nucleic acid sequence) without appreciable loss of biological utility or activity and possibly with an increase in such utility or activity.

The present invention further provides antibodies which specifically bind the DRM polypeptide. The antibodies of the present invention include both polyclonal and monoclonal antibodies. Such antibodies may be murine, fully human, chimeric or humanized. These antibodies can also include Fab or $F(ab')_2$ fragments, as well as single chain antibodies (ScFv) (90). The antibodies can be of any isotype IgG, IgA,

IgD, IgE and IgM. The antibodies can be produced against peptides which are identified to be immunogenic peptides as described in the Examples provided herein and according to methods well known in the art for identifying immunogenic regions in an amino acid sequence. Such antibodies can be produced by techniques well known in the art which include those described in Kohler et al. (42) or U.S. Patents 5,545,806, 5,569,825 and 5,625,126, incorporated herein by reference.

The antibodies of this invention can be used to detect and/or quantitate DRM in a sample. For example, a method is provided for detecting and/or quantitating a DRM protein or antigen in a sample, which can be a biological sample, comprising contacting the sample with an antibody which specifically binds DRM under conditions whereby an antigen/antibody complex can form and detecting the presence of the complex, whereby the presence of the antigen/antibody complex indicates the presence of a DRM protein or antigen in the sample. The amount of the DRM protein in the detected antigen/antibody complex can be determined by methods well known in the art for quantitating protein.

Conditions whereby an antigen/antibody complex can form as well as assays for the detection of the formation of an antigen/antibody complex and quantitating of the detected protein are standard in the art. Such assays can include, but are limited to, Western blotting, immunoprecipitation, immunofluorescence, immunocytochemistry, immunohistochemistry, fluorescence activated cell sorting (FACS), immunomagnetic assays, ELISA, agglutination assays, flocculation assays, cell panning, etc., as are well known to the artisan.

The DRM protein of the present invention has been identified to play a role in regulating a cell's proliferation cycle, as set forth in the Examples provided herein. Thus, the DRM protein of this invention and nucleic acids encoding DRM have therapeutic utility in applications in which it is desirable to alter or control a cell's proliferation cycle.

In particular, the present invention provides a method of arresting cell growth, comprising administering to the cell an effective amount of DRM protein or active fragment thereof. The cell can be *in vivo* or *ex vivo* and the DRM protein or active fragment thereof can be in a pharmaceutically acceptable carrier. As used herein, an "active fragment thereof" is a fragment of DRM identified to possess the cell growth arresting activity of the complete protein. Such an active fragment can be identified by producing fragments of the DRM proteins according to standard protocols and assaying the fragments for cell growth arresting activity according to the methods described herein. Also as used herein, "arresting cell growth" means treating or modifying the cell such that the cell is unable to proliferate or form colonies when plated on tissue culture dishes in appropriate media under conditions where similar untreated or unmodified cells, but otherwise identical cells will do so. An effective amount of DRM or active fragment thereof is that amount which results in arrest of cell growth as measured by labeling index, presence of mitotic figures or any other cell proliferation assay now known or developed in the future.

Furthermore, the present invention provides a method of treating or preventing a hyperproliferative cell disorder in a subject diagnosed with, or at risk of developing, a hyperproliferative cell disorder, comprising administering to the subject an effective amount of DRM protein or an active fragment thereof, in a pharmaceutically acceptable carrier. As used herein, an "active fragment thereof" is a fragment of DRM identified to possess the hyperproliferative cell disorder treating or preventing activity of the complete protein. Such an active fragment can be identified by producing fragments of the DRM proteins according to standard protocols and assaying the fragments for hyperproliferative cell disorder treating or preventing activity according to the methods described herein.

The subject can be any animal in which DRM can function in regulating the growth of a cell and can treat or prevent a hyperproliferative cell disorder. For example, the subject can be a mammal and is most preferably a human. As used herein, a "hyperproliferative cell disorder" is any disorder of a cell characterized by unregulated cell division and growth and which has a deleterious effect. An example of

a hyperproliferative cell disorder is cancer. Thus, the DRM protein or active fragment thereof of the present invention can be administered to a subject diagnosed with a cancer, to treat the subject's cancer. Examples of cancers include, but are not limited to, leukemia, lymphoma, myeloma, melanoma, sarcoma, bone cancer, prostate cancer, lung cancer, renal cancer, etc.

As stated above, the DRM protein of the present invention can be in a pharmaceutically acceptable carrier and in addition, can include other medicinal agents, pharmaceutical agents, carriers, adjuvants, diluents, immunostimulatory cytokines, etc. By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to an individual along with the DRM protein without causing substantial deleterious biological effects or interacting in a deleterious manner with any of the other components of the composition in which it is contained. Actual methods of preparing such dosage forms are known, or will be apparent, to those skilled in this art; for example, see *Remington's Pharmaceutical Sciences* (91).

To determine the effect of the administration of the DRM polypeptide or active fragment thereof on inhibition of tumor cell growth in laboratory animals, the animals can either be pre-treated with the DRM polypeptide or active fragment thereof and then challenged with a lethal dose of tumor cells, or the lethal dose of tumor cells can be administered to the animal prior to receipt of the DRM polypeptide or active fragment thereof and survival times documented. To determine the amount of DRM polypeptide or active fragment thereof which would be an effective tumor cell growth-inhibiting amount, animals can be treated with tumor cells as described herein and varying amounts of the DRM polypeptide or active fragment thereof can be administered to the animals. Standard clinical parameters, as described herein, can be measured and that amount of DRM polypeptide or active fragment thereof effective in inhibiting tumor cell growth can be determined. These parameters, as would be known to one of ordinary skill in the art of oncology and tumor biology, can include, but are not limited to, physical examination of the subject, measurements of tumor size, measurements of levels of circulating tumor antigen, X-ray studies and biopsies, as well as any other

assay now known or later identified as a diagnostic and/or prognostic assay for tumor cell growth.

In vitro assays can also be utilized to determine the effect of the administration of the DRM polypeptide or active fragment thereof on inhibition of tumor cell growth. These assays are well known in the art and include *in vitro* invasiveness assays.

Once dosages effective in treating hyperproliferative cell disorders, such as cancer, are determined for animal models, these data can be extrapolated to determine approximate effective treatment dosages in humans (e.g., by correlating mg/kg body weight of an amount of DRM protein effective in animals). Specific effective hyperproliferative cell disorder treating dosages in humans can be determined according to standard protocols established for clinical trials, as are well documented in the art (45-49). To determine the efficacy of administration of a given dose of the DRM polypeptide or active fragment thereof for treating hyperproliferative cell disorders, such as cancer, in humans, standard clinical response parameters can be analyzed, as described herein and as are well known in the art.

Additionally, the efficacy of administration of a particular dose of DRM protein or active fragment thereof in preventing a hyperproliferative cell disorder, such as cancer, in a subject not known to have a hyperproliferative cell disorder, but known to be at risk of developing a hyperproliferative cell disorder, can be determined by evaluating standard signs, symptoms and objective laboratory tests, known to one of skill in the art, over time after administration of the DRM polypeptide or active fragment thereof. This time interval may be short (weeks/months) or long (years/decades). The determination of who would be at risk for the development of a hyperproliferative cell disorder would be made based on current knowledge of the known risk factors for a particular disorder familiar to clinicians and researchers in this field, such as a particularly strong family history of a disorder. Furthermore, a subject can be identified as being at risk of developing a hyperproliferative disorder, such as cancer, according to the methods provided herein.

The DRM polypeptide or active fragment thereof of this invention can be administered to the subject orally or parenterally, as for example, by intramuscular injection, by intraperitoneal injection, topically, transdermally, injection directly into the tumor, or the like, although subcutaneous injection is typically preferred. Tumor cell growth inhibiting and cancer treating amounts of the DRM polypeptide or active fragment thereof can be determined using standard procedures, as described. The exact dosage of the DRM polypeptide or active fragment thereof will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the cancer or disorder that is being treated, the mode of administration and the like. Thus, it is not possible to specify an exact amount. However, an appropriate amount may be determined by one of ordinary skill in the art using only routine screening given the teachings herein.

For oral administration, fine powders or granules may contain diluting, dispersing, and/or surface active agents and may be presented in water or in a syrup, in capsules or sachets in the dry state, or in a nonaqueous solution or suspension wherein suspending agents may be included, in tablets wherein binders and lubricants may be included, or in a suspension in water or a syrup. Where desirable or necessary, flavoring, preserving, suspending, thickening, or emulsifying agents may be included. Tablets and granules are preferred oral administration forms and these may be coated.

Parenteral administration, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system, such that a constant dosage level is maintained. See, e.g., U.S. Patent No. 3,710,795, which is incorporated by reference herein.

For solid compositions, conventional nontoxic solid carriers include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talc, cellulose, glucose, sucrose, magnesium carbonate, and the like.

Liquid pharmaceutically administrable compositions can, for example, be prepared by dissolving, dispersing, etc. an active compound as described herein and optional pharmaceutical adjuvants in an excipient, such as, for example, water, saline, aqueous dextrose, glycerol, ethanol, and the like, to thereby form a solution or suspension. If desired, the pharmaceutical composition to be administered may also contain minor amounts of nontoxic auxiliary substances such as wetting or emulsifying agents, pH buffering agents and the like, for example, sodium acetate, sorbitan monolaurate, triethanolamine sodium acetate, triethanolamine oleate, etc. Actual methods of preparing such dosage forms are known, or will be apparent, to those skilled in this art (91).

Generally, to treat or prevent a hyperproliferative cell disorder in a subject, the dosage of DRM protein or active fragment thereof will approximate that which is typical for the administration of proteins and typically, the dosage will be in the range of about 1 to 500 μg of the DRM polypeptide or active fragment thereof per dose, and preferably in the range of 50 to 250 μg of the DRM polypeptide or active fragment thereof per dose. This amount can be administered to the subject once every other week for about eight weeks or once every other month for about six months. The effects of the administration of the DRM polypeptide or active fragment thereof can be determined starting within the first month following the initial administration and continued thereafter at regular intervals, as needed, for an indefinite period of time.

As described herein, the present invention also provides a nucleic acid and a vector, which can be in a pharmaceutically acceptable carrier, which encodes the DRM polypeptide or active fragments thereof, of the present invention. Such nucleic acids can be used in gene therapy protocols to treat or prevent hyperproliferative cell disorders, such as a cancer, in a subject.

Thus, the present invention further provides a method of treating a hyperproliferative cell disorder in a subject diagnosed with a hyperproliferative cell disorder, comprising administering an effective amount of the nucleic acid of this invention, which encodes the DRM protein or an active fragment thereof, to a cell of

the subject under conditions whereby the nucleic acid is expressed in the subject's cell, thereby treating the hyperproliferative cell disorder.

Also provided is a method of arresting the growth of a cell, comprising administering to the cell an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof, to a cell under conditions whereby the nucleic acid is expressed in the cell, thereby arresting the growth of the cell.

The present invention further provides a method of inhibiting tumor cell growth, comprising administering to a tumor cell an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof, to a tumor cell under conditions whereby the nucleic acid is expressed in the tumor cell, thereby inhibiting tumor cell growth.

The nucleic acid can be administered to the cell in a virus, which can be, for example, adenovirus, retrovirus and adeno-associated virus. Alternatively, the nucleic acid of this invention can be administered to the cell as naked DNA or in a liposome. The cell can be either *in vivo* or *ex vivo*. Also, the cell can be any cell which can take up and express exogenous nucleic acid and produce the DRM polypeptide or fragment thereof of this invention.

If *ex vivo* methods are employed, cells or tissues can be removed and maintained outside the subject's body according to standard protocols well known in the art. The nucleic acids of this invention can be introduced into the cells via any gene transfer mechanism, such as, for example, virus-mediated gene delivery, calcium phosphate mediated gene delivery, electroporation, microinjection or proteoliposomes. The transduced cells can then be infused (e.g., in a pharmaceutically acceptable carrier) or transplanted back into the subject per standard methods for the cell or tissue type. Methods for transplantation or infusion of various cells into a subject are well known in the art.

For *in vivo* methods, the nucleic acid encoding the DRM protein or active fragments thereof, can be administered to the subject in a pharmaceutically acceptable carrier as further described herein.

In the methods described above which include the administration and uptake of exogenous nucleic acid into the cells of a subject (i.e., gene transduction or transfection), the nucleic acids of the present invention can be in the form of naked nucleic acid or the nucleic acids can be in a vector for delivering the nucleic acids to the cells for expression of the DRM protein or active fragment thereof. The vector can be a commercially available preparation, such as an adenovirus vector (Quantum Biotechnologies, Inc. (Laval, Quebec, Canada). Delivery of the nucleic acid or vector to cells can be via a variety of mechanisms. As one example, delivery can be via a liposome, using commercially available liposome preparations such as LIPOFECTIN, LIPOFECTAMINE (GIBCO-BRL, Inc., Gaithersburg, MD), SUPERFECT (Qiagen, Inc. Hilden, Germany) and TRANSFECTAM (Promega Biotec, Inc., Madison, WI), as well as other liposomes developed according to procedures standard in the art. In addition, the nucleic acid or vector of this invention can be delivered *in vivo* by electroporation, the technology for which is available from Genetronics, Inc. (San Diego, CA) as well as by means of a SONOPORATION machine (ImaRx Pharmaceutical Corp., Tucson, AZ).

As one example, vector delivery can be via a viral system, such as a retroviral vector system which can package a recombinant retroviral genome (see e.g.,50,51). The recombinant retrovirus can then be used to infect and thereby deliver to the infected cells nucleic acid encoding the DRM protein. The exact method of introducing the exogenous nucleic acid into mammalian cells is, of course, not limited to the use of retroviral vectors. Other techniques are widely available for this procedure including the use of adenoviral vectors (52), adeno-associated viral (AAV) vectors (53), lentiviral vectors (54), pseudotyped retroviral vectors (55). Physical transduction techniques can also be used, such as liposome delivery and receptor-mediated and other endocytosis mechanisms (see, for example, 56). This invention can be used in conjunction with any of these or other commonly used gene transfer methods.

Various adenoviruses may be used in the compositions and methods described herein. For example, a nucleic acid encoding the DRM protein can be inserted within the genome of adenovirus type 5. Similarly, other types of adenovirus may be used such as type 1, type 2, etc. For an exemplary list of the adenoviruses known to be able to infect human cells and which therefore can be used in the present invention, see Fields, *et al.* (57). Furthermore, it is contemplated that a recombinant nucleic acid comprising an adenoviral nucleic acid from one type adenovirus can be packaged using capsid proteins from a different type adenovirus.

The adenovirus of the present invention is preferably rendered replication deficient, depending upon the specific application of the compounds and methods described herein. Methods of rendering an adenovirus replication deficient are well known in the art. For example, mutations such as point mutations, deletions, insertions and combinations thereof, can be directed toward a specific adenoviral gene or genes, such as the E1 gene. For a specific example of the generation of a replication deficient adenovirus for use in gene therapy, see WO 94/28938 (Adenovirus Vectors for Gene Therapy Sponsorship) which is incorporated herein.

In the present invention, the nucleic acid encoding the DRM protein or active fragment thereof (DRM-encoding insert) can be inserted within an adenoviral genome and the DRM-encoding insert can be positioned such that an adenovirus promoter is operatively linked to the DRM-encoding insert such that the adenoviral promoter can then direct transcription of the nucleic acid, or the DRM-encoding insert may contain its own adenoviral promoter. Similarly, the DRM-encoding insert may be positioned such that the nucleic acid encoding the DRM protein or fragment may use other adenoviral regulatory regions or sites such as splice junctions and polyadenylation signals and/or sites. Alternatively, the nucleic acid encoding the DRM protein or fragment may contain a different enhancer/promoter (e.g., CMV or RSV-LTR enhancer/promoter sequences) or other regulatory sequences, such as splice sites and polyadenylation sequences, such that the nucleic acid encoding the DRM protein or fragment may contain those sequences necessary for expression of the DRM protein fragment and not partially or totally require these regulatory regions and/or sites of the

adenovirus genome. These regulatory sites may also be derived from another source, such as a virus other than adenovirus. For example, a polyadenylation signal from SV40 or BGH may be used rather than an adenovirus, a human, or a murine polyadenylation signal. The DRM-encoding insert may, alternatively, contain some sequences necessary for expression of the nucleic acid encoding the DRM protein or fragment and derive other sequences necessary for the expression of the DRM-encoding insert from the adenovirus genome, or even from the host in which the recombinant adenovirus is introduced.

As another example, for administration of nucleic acid encoding the DRM protein or active fragment thereof to an individual in an AAV vector, the AAV particle can be directly injected intravenously. The AAV has a broad host range, so the vector can be used to transduce any of several cell types, but preferably cells in those organs that are well perfused with blood vessels. To more specifically administer the vector, the AAV particle can be directly injected into a target organ, such as muscle, liver or kidney. Furthermore, the vector can be administered intraarterially, directly into a body cavity, such as intraperitoneally, or directly into the central nervous system (CNS).

An AAV vector can also be administered in gene therapy procedures in various other formulations in which the vector plasmid is administered after incorporation into other delivery systems such as liposomes or systems designed to target cells by receptor-mediated or other endocytosis procedures. The AAV vector can also be incorporated into an adenovirus, retrovirus or other virus which can be used as the delivery vehicle.

As described above, the nucleic acid or vector of the present invention can be administered *in vivo* in a pharmaceutically acceptable carrier. By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to a subject, along with the nucleic acid or vector, without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained. The carrier would naturally be selected to minimize any degradation of

the active ingredient and to minimize any adverse side effects in the subject, as would be well known to one of skill in the art.

The mode of administration of the nucleic acid or vector of the present invention can vary predictably according to the disorder being treated and the tissue being targeted. For example, for administration of the nucleic acid or vector in a liposome, catheterization of an artery upstream from the target organ is a preferred mode of delivery, because it avoids significant clearance of the liposome by the lung and liver.

The nucleic acid or vector may be administered orally, parenterally (e.g., intravenously), by intramuscular injection, by intraperitoneal injection, transdermally, extracorporeally, topically or the like, although intravenous administration is typically preferred. The exact amount of the nucleic acid or vector required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the disorder being treated, the particular nucleic acid or vector used, its mode of administration and the like. Thus, it is not possible to specify an exact amount for every nucleic acid or vector. However, an appropriate amount can be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein (91).

As one example, if the nucleic acid of this invention is delivered to the cells of a subject in an adenovirus vector, the dosage for administration of adenovirus to humans can range from about 10^7 to 10^9 plaque forming units (pfu) per injection, but can be as high as 10^{12} pfu per injection (59,60). Ideally, a subject will receive a single injection. If additional injections are necessary, they can be repeated at six month intervals for an indefinite period and/or until the efficacy of the treatment has been established.

Parenteral administration of the nucleic acid or vector of the present invention, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution of suspension in liquid prior to injection, or as emulsions. A more recently

revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant dosage is maintained. See, e.g., U.S. Patent No. 3,610,795, which is incorporated by reference herein.

To determine the effect of the administration of the nucleic acid of this invention on inhibition of tumor cell growth in laboratory animals, the animals can either be pre-treated with the nucleic acid and then challenged with a lethal dose of tumor cells, or the lethal dose of tumor cells can be administered to the animal prior to receipt of the nucleic acid and survival times documented. To determine the amount of nucleic acid which would be an effective tumor cell growth-inhibiting amount, animals can be treated with tumor cells as described herein and varying amounts of the nucleic acid can be administered to the animals. Standard clinical parameters, as described herein, can be measured and the amount of DRM encoding nucleic acid effective in inhibiting tumor cell growth can be determined. These parameters, as would be known to one of ordinary skill in the art of oncology and tumor biology, can include, but are not limited to, physical examination of the subject, measurements of tumor size, measurements of levels of circulating tumor antigen, X-ray studies and biopsies, as well as any other assay now known or later identified as a diagnostic and/or prognostic assay for tumor cell growth.

Once dosages effective in inhibiting cell growth and/or treating hyperproliferative cell disorders, such as cancer, are determined for animal models, these data can be extrapolated to determine approximate effective treatment dosages in humans. Specific effective hyperproliferative cell disorder treating dosages of DRM-encoding DNA in humans can be determined according to standard protocols established for clinical trials, as are well documented in the art. To determine the efficacy of administration of a given dose of the DRM-encoding nucleic acid for treating hyperproliferative cell disorders, such as cancer, in humans, standard clinical response parameters can be analyzed, as described herein and as are well known in the art.

Additionally, the efficacy of administration of a particular dose of DRM encoding nucleic acid in preventing a hyperproliferative cell disorder, such as cancer, in a subject not known to have a hyperproliferative cell disorder, but known to be at risk of developing a hyperproliferative cell disorder, can be determined by evaluating standard signs, symptoms and objective laboratory tests, known to one of skill in the art, over time after administration of the DRM encoding nucleic acid. This time interval may be short (weeks/months) or long (years/decades). The determination of who would be at risk for the development of a hyperproliferative cell disorder would be made based on current knowledge of the known risk factors for a particular disorder familiar to clinicians and researchers in this field, such as a particularly strong family history of a disorder. Furthermore, a subject can be identified as being at risk of developing a hyperproliferative disorder, such as cancer, according to the methods provided herein.

As described herein, the DRM protein is produced in normal cells (i.e., cells which are differentiating normally) at detectable levels. Tumor cells and cells which have been transformed by transfection with an oncogene do not produce detectable levels of DRM protein. A decrease in the level of DRM protein or RNA, or such a decrease in a particular differentiating lineage which normally expresses DRM during differentiation, can be diagnostic of a premalignant or early malignant state. Thus, the present invention provides a method for the early identification of malignancies or premalignant states.

Thus, further provided in the present invention is a method of identifying a subject at risk of developing a hyperproliferative cell disorder (e.g., cancer), comprising measuring the amount of DRM protein or the amount of nucleic acid encoding DRM in a cell of the subject, whereby an amount of DRM protein or nucleic acid encoding DRM in a cell less than the amount of DRM protein or nucleic acid encoding DRM in a cell of a normal subject identifies a subject at risk of developing a hyperproliferative cell disorder. The cell of the subject is a cell which produces DRM and can be, but is not limited to cells of the brain, lung, intestine and esophagus (goblet cells), as well as any other cell now known or later identified to produce DRM.

The amount of DRM protein in a cell can be determined by methods standard in the art for quantitating proteins in a cell, such as Western blotting, ELISA, ELISPOT, immunoprecipitation, immunofluorescence (e.g., FACS), immunohistochemistry, immunocytochemistry, etc., as well as any other method now known or later developed for quantitating protein in a cell.

The amount of nucleic acid encoding DRM in a cell can be determined by methods standard in the art for quantitating nucleic acid in a cell, such as *in situ* hybridization, quantitative PCR, Northern blotting, ELISPOT, dot blotting, etc., as well as any other method now known or later developed for quantitating nucleic acid in a cell.

The cell can be a separate cell or a cell in intact tissue, which can be a biopsy specimen. As used herein, "a cell of a normal subject" means a cell or tissue which is histologically normal and was obtained from a subject believed to be without malignancy and having no increased rick of developing a malignancy or was obtained from tissues adjacent to tissue known to be malignant and which is determined to be histologically normal (non-malignant) as determined by a pathologist.

The present invention is further based on the unexpected discovery that fusion of DRM or active fragments thereof, with enhanced green fluorescent protein (EGFP) or active fragments thereof, yields a protein which is localized to the nucleus, rather than the cytoplasm, and results in an improved EGFP which has greater stability than conventional EGFP, providing a much more versatile research tool for use in screening assays, protein-protein interaction studies and cell marking applications.

Thus, the present invention provides a fusion polypeptide comprising a DRM protein region and a green fluorescent protein region. For example, the fusion polypeptide of this invention can be a polypeptide having the amino acid sequence of SEQ ID NO:29. The fusion polypeptide of this invention can comprise the entire DRM protein or an active fragment thereof and the entire EGFP or an active fragment thereof. The identification of an active fragment of either DRM or EGFP can be carried out

according to routine methods for identifying active fragments. For example, a fragment of either protein can be produced by PCR amplification of a specific region of the protein, by deleting portions of the protein at specific restriction sites with restriction endonucleases, by introducing stop codons into the protein sequence, by synthesizing a peptide comprising a fragment of the protein, etc., as would be well known to one of skill in the art. The resulting fragments can be tested for functional activity according to the methods provided herein as well as are described in the art. For example, the fusion protein of this invention can have the amino acid sequence of SEQ ID NOS:30, 31, 32, 33, 34 and 35, encoded by the nucleic acids of SEQ ID NOS:5, 6, 7, 8, 9 and 19, respectively. The production of each of the fusion proteins having the amino acid sequences of SEQ ID NOS:30-35 is described in the Examples section herein.

The present invention further provides a green fluorescent protein having increased stability, comprising a fusion protein comprising a DRM protein amino acid sequence linked to an EGFP amino acid sequence. As used herein, "having increased stability" means that the EGFP of the EGFP/DRM fusion protein maintains fluorescence activity when exposed to fixatives (e.g., ethanol, methanol, acetone), detergents (e.g., TritonX100, NP40), or other conditions under which the fluorescence activity of unfused (conventional) EGFP is greatly diminished (>75%) or no longer detectable.

An isolated nucleic acid encoding the fusion polypeptides described above is also provided. The isolated nucleic acid of this invention which encodes the EGFP/DRM fusion protein can be a nucleic acid having the nucleotide sequence of SEQ ID NO:1. By "isolated nucleic acid" is meant a nucleic acid molecule that is substantially free of the other nucleic acids and other components commonly found in association with nucleic acid in a cellular environment. Separation techniques for isolating nucleic acids from cells are well known in the art and include phenol extraction followed by ethanol precipitation and rapid solubilization of cells by organic solvent or detergents (35).

The nucleic acid encoding the fusion polypeptide can be any nucleic acid that functionally encodes the fusion polypeptide. To functionally encode the polypeptide (i.e., allow the nucleic acid to be expressed), the nucleic acid can include, for example, expression control sequences, such as an origin of replication, a promoter, an enhancer and necessary information processing sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites and transcriptional terminator sequences. Preferred expression control sequences are promoters derived from metallothionine genes, actin genes, immunoglobulin genes, CMV, SV40, adenovirus, bovine papilloma virus, etc. A nucleic acid encoding a selected fusion polypeptide can readily be determined based upon the genetic code for the amino acid sequence of the selected fusion polypeptide and many nucleic acids will encode any selected fusion polypeptide. Modifications in the nucleic acid sequence encoding the fusion polypeptide are also contemplated. Modifications that can be useful are modifications to the sequences controlling expression of the fusion polypeptide to make production of the fusion polypeptide inducible or repressible as controlled by the appropriate inducer or repressor. Such means are standard in the art (*35*). The nucleic acids can be generated by means standard in the art, such as by recombinant nucleic acid techniques, as exemplified in the examples herein and by synthetic nucleic acid synthesis or *in vitro* enzymatic synthesis.

A vector comprising the nucleic acids encoding the fusion proteins of the present invention and a cell comprising the vector are also provided. The vector can be in a host (e.g., cell line or transgenic animal) that can express the fusion polypeptide contemplated by the present invention.

There are numerous *E. coli* (*Escherichia coli*) expression systems known to one of ordinary skill in the art useful for the expression of nucleic acid encoding proteins such as fusion proteins. Other microbial hosts suitable for use include bacilli, such as *Bacillus subtilis*, and other enterobacteria, such as *Salmonella* and *Serratia*, as well as various *Pseudomonas* species. These prokaryotic hosts can support expression vectors which will typically contain expression control sequences compatible with the host cell (e.g., an origin of replication). In addition, any number of a variety of well-known

promoters will be present, such as the lactose promoter system, a tryptophan (Trp) promoter system, a beta-lactamase promoter system, or a promoter system from phage lambda. The promoters will typically control expression, optionally with an operator sequence and have ribosome binding site sequences for example, for initiating and completing transcription and translation. If necessary, an amino terminal methionine can be provided by insertion of a Met codon 5' and in-frame with the protein sequences. Also, the carboxy-terminal extension of the protein can be removed using standard oligonucleotide mutagenesis procedures.

Additionally, yeast expression can be used. There are several advantages to yeast expression systems. First, evidence exists that proteins produced in a yeast secretion system exhibit correct disulfide pairing. Second, post-translational glycosylation is efficiently carried out by yeast secretory systems. The *Saccharomyces cerevisiae* pre-pro-alpha-factor leader region (encoded by the *MFα-1* gene) is routinely used to direct protein secretion from yeast (89). The leader region of pre-pro-alpha-factor contains a signal peptide and a pro-segment which includes a recognition sequence for a yeast protease encoded by the *KEX2* gene. This enzyme cleaves the precursor protein on the carboxyl side of a Lys-Arg dipeptide cleavage-signal sequence. The polypeptide coding sequence can be fused in-frame to the pre-pro-alpha-factor leader region. This construct is then put under the control of a strong transcription promoter, such as the alcohol dehydrogenase I promoter or a glycolytic promoter. The protein coding sequence is followed by a translation termination codon, which is followed by transcription termination signals. Alternatively, the polypeptide coding sequence of interest can be fused to a second protein coding sequence, such as Sj26 or β-galactosidase, used to facilitate purification of the fusion protein by affinity chromatography. The insertion of protease cleavage sites to separate the components of the fusion protein is applicable to constructs used for expression in yeast.

Efficient post-translational glycosylation and expression of recombinant proteins can also be achieved in *Baculovirus* systems in insect cells.

Mammalian cells permit the expression of proteins in an environment that favors important post-translational modifications such as folding and cysteine pairing, addition of complex carbohydrate structures and secretion of active protein. Vectors useful for the expression of proteins in mammalian cells are characterized by insertion of the protein coding sequence between a strong viral promoter and a polyadenylation signal. The vectors can contain genes conferring either gentamicin or methotrexate resistance for use as selectable markers. The fusion protein coding sequence can be introduced into a Chinese hamster ovary (CHO) cell line using a methotrexate resistance-encoding vector. Presence of the vector RNA in transformed cells can be confirmed by Northern blot analysis and production of a cDNA or opposite strand RNA corresponding to the fusion protein coding sequence can be confirmed by Southern and Northern blot analysis, respectively. A number of other suitable host cell lines capable of secreting intact proteins have been developed in the art and include the CHO cell lines, HeLa cells, myeloma cell lines, Jurkat cells and the like. Expression vectors for these cells can include expression control sequences, as described above.

The vectors containing the nucleic acid sequences of interest can be transferred into the host cell by well-known methods, which vary depending on the type of cell host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment or electroporation may be used for other cell hosts.

Alternative vectors for the expression of protein in mammalian cells, similar to those developed for the expression of human gamma-interferon, tissue plasminogen activator, clotting Factor VIII, hepatitis B virus surface antigen, protease Nexinl, and eosinophil major basic protein, can be employed. Further, the vector can include CMV promoter sequences and a polyadenylation signal available for expression of inserted nucleic acid in mammalian cells (such as COS7).

The nucleic acid sequences can be expressed in hosts after the sequences have been positioned to ensure the functioning of an expression control sequence. These expression vectors are typically replicable in the host organisms either as episomes or

as an integral part of the host chromosomal DNA. Commonly, expression vectors can contain selection markers, e.g., tetracycline resistance or hygromycin resistance, to permit detection and/or selection of those cells transformed with the desired nucleic acid sequences (*see*, e.g., U.S. Patent 4,704,362).

Thus, further provided is a method of producing the green fluorescent protein having increased stability of this invention, comprising the steps of producing a nucleic acid construct whereby a first nucleic acid sequence encoding EGFP or an active fragment thereof is positioned upstream and in frame with a second nucleic acid encoding DRM or an active fragment thereof; cloning the nucleic acid construct into an expression vector; and placing the expression vector into a cell under conditions whereby the nucleic acid of the construct will be expressed, thereby producing a green fluorescent protein having increased stability. The expression vector and expression system can be of any of the types as described herein. The cloning of the first and second nucleic acids into the expression vector and expression of the nucleic acids under conditions which allow for the production of the fusion protein of this invention can be carried out as described in the Examples section included herein. The method of this invention can further comprise the step of isolating and purifying the fusion polypeptide, according to methods well known in the art and as described herein.

The EGFP/DRM fusion protein of this invention improves the stability of the EGFP as compared to conventional EGFP.. Thus, the fusion protein of this invention can be used in assays for which conventional EGFP is not suitable, such as fluorescence-based assays which require cell fixation and in protocols where cell marking is necessary or desired. For example, the EGFP/DRM fusion protein of this invention can be used in cell cycle analysis using PI or BudR, where fixation is required to allow the dye to enter in to the cell nucleus. Also, the stabilized EGFP of this invention can be introduced as a marker (e.g., linked to a ligand to detect the presence of a receptor) or the nucleic acid encoding the stabilized EGFP can be used to identify cells into which a particular expression construct is introduced or where a reporter gene signal is desired.

The stabilized EGFP of this invention can also be linked to proteins or antibodies for use in ELISA assays. The advantage of using stabilized EGFP is that the stabilized EGFP can be attached as a particular protein is being synthesized, so that materials which could not be chemically modified to attach fluorescent groups because of stability problems could be labeled. The stabilized EGFP can also be used as a marker during purification. For example, materials can be produced *in vivo* in fermentor-type production facilities and a desired material can be purified by the presence of the EGFP protein marker.

The present invention is more particularly described in the following examples which are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

EXAMPLES

Example I. Isolation and characterization of rat *drm* gene and gene product.

Cell culture. The REF-1, DTM, F-1 and ST33c rat cell lines have previously been described (40-42). DTM and ST33c cell lines were maintained at 34°C in DMEM with 5% fetal calf serum, while REF-1, as well as REF-1cells transformed by different oncogenes, were grown at 37°C in DMEM (Gibco) with 5% or 10% fetal calf serum.

DNA and RNA analysis. High molecular weight DNA was purified by standard procedures (15) and analyzed by Southern blotting (35). Total RNA was extracted from culture cells by RNAzolB (Tel-Teck, Inc., Texas) (7), and 10 μg was used per lane in a Northern analysis. Filters were pre-hybridized and hybridized at 42°C for 18-20 hr in 5 x SSPE (NaCl, NaH_2PO_4, Na_2EDTA, pH 7.4) containing 10X Denhardt's solution (9), 2% SDS, 50% formamide, and 100 μg of heat-denatured salmon sperm DNA per ml, the filters were washed sequentially in 2 x SSC/0.05% SDS at room temperature for 30 min and in 0.1 x SSC/0.1% SDS at 50°C for 40 min. Autoradiography was for 2-4 days at -70°C with an intensifying screen. Poly(A)$^+$ was isolated by using the "Fast Track" mRNA isolation kit (InVitrogen) according to the manufacturer's specifications. Multi-tissue Northern blot (Clontech) was treated according to the manufacturer's protocol.

The murine recombinant retrovirus expressing v-*src* was obtained from S. M. Anderson. The vector expressing activated *ras* is pEJ-*ras* (38) containing the Val^{12}-mutated fragments of human c-*ras* in pBR322.

Identification and isolation of *drm* cDNA.

Messenger RNAs expressed differentially in DTM and F-1 cells were displayed as described by Liang and Pardee (25). First-strand cDNAs were synthesized on 1.5 μg of polyadenylated RNA extracted from either cell line using the "cDNA Cycle Kit for RT-PCR" (Invitrogen) and specific primers T12VA, T12VC (V was either A, C, G). cDNAs were then amplified by polymerase chain reaction (PCR) using [α-^{35}S]dATP and combinations of 3' specific primers and arbitrary 5' primers [AGCCAGCGAA (SEQ ID NO:22), GACCGCTTGT (SEQ ID NO:23), AGGTGACCGT (SEQ ID NO:24), GGTACTCCAC (SEQ ID NO:25), GTTGCGATCC (SEQ ID NO:26)]. PCR products were separated on a 6% polyacrylamide gel and visualized by autoradiography.

Screening of cDNA library. An oligo dT-primed cDNA library of rat embryo fibroblasts constructed in a λZAP XR vector, was screened with the 691 bp *drm* cDNA isolated from F-1 mRNA by the differential display technique, as described (35). Three independent clones (Cl3ZAP, C17ZAP and Cl10ZAP) were isolated and further analyzed. 5' sequences of the Cl7ZAP absent from the other clones were used as probes to screen a rat kidney 5'-stretch λgt11 cDNA library (Clontech). Two clones (Cl7gt, Cl10gt) were isolated, further amplified and analyzed. cDNA clones were sequenced on both strands by the dideoxy chain termination method using the "T7 sequencing kit" (Pharmacia Biotech) (36). Portions of the sequencing data were compiled and analyzed by using the University of Wisconsin Genetics Computer Group package (11).

Protein analysis.

1) *In vitro* transcription and translation. The 2.1 kb EcoRI fragment of Clone 10 gt, as well as the BamHI/KpnI fragment from this insert, both containing the putative *drm* coding region, were inserted into the Bluescript KS vector. Plasmid

DNAs were transcribed and translated using the TNT T7 and T3 reticulocyte lysate system (Promega) with L-^{35}S-cysteine (1200 Ci/mmol, Amersham). Translation products were separated by SDS-PAGE and processed for fluorography. T7 polymerase produces a sense message, while T3 produces an antisense product. Luciferase DNA was used as a positive control.

2) Construction of tagged *drm* protein-expression vector. The coding region of *drm* cDNA was fused in frame at its 3' end with the DNA fragment encoding the nine residue epitope of the HA-1 influenza virus hemagglutinin by polymerase chain reaction. The primers used were: 5' (5'-CCGCTCGAGGTGACAGAATGAATCGC-3') (SEQ ID NO:27) and 3' (5'CCCGTTAACTTAGGCGTAGTCGGGCACGTCGTAGGGGTAATCCAAGTCG AT3') (SEQ ID NO:28). The 5' primer introduces an XhoI restriction site, while the 3' primer removes the stop codon from the *drm* and introduces another one downstream from the inserted HA-1 sequence. It also introduces an HpaI site downstream from the stop codon. The PCR product was digested with XhoI/HpaI and inserted into the pSVL expression vector (39) between the XhoI and SmaI sites.

3) Preparation and characterization of antibodies. Two peptides based on the predicted sequence of drm protein were selected to raise rabbit polyclonal antibodies. An N-terminal cysteine residue was added to the first peptide (990), which corresponds to amino acids 79-92 to enable coupling of the peptide to KLH (keyhole limpet hemocyanim) carrier protein prior to immunization. The second peptide (987), corresponding to amino acids 158-172, was coupled to the carrier protein through a natural cysteine residue on its N-terminal end. A peptide which corresponds to amino acids 33-52 was expressed as a fusion with bacteriophage MS2 coat protein and used to immunize rabbits as described herein.

4) Immunoprecipitation and Western blotting. Cell lysates prepared under denaturing conditions were either first immunoprecipitated using either *drm*-specific 990 antisera or anti-HA monoclonal antibody (Babco), followed by separation on SDS-

PAGE and Western blotting, or total lysates were analyzed by SDS-PAGE and Western blotting.

For immunoblotting, proteins were electrophoretically transferred to nitrocellulose at 60 mA for 2 hrs. Filters were incubated first with the appropriate primary antibody and then with horseradish peroxidase-labeled secondary antibodies (Amersham). Antibodies were detected using the ECL detection system (Amersham) or the Super Signal CL-HRP Substrate System (Pierce) and visualized using Kodak XAR-5 X-ray film.

Western blots were "stripped" for reprobing with other primary antibodies according to the manufacturer's protocol (Amersham).

Transfection of *drm* expression vectors. For stable transfection experiments, cDNA containing the full-length *drm* ORF was inserted into the BamHI and KpnI restriction sites of the pMEXneo expression vector (21). In this construct, *drm* and the *neo*-selectable marker were under the control of an MuLV LTR and an SV40 promoter, respectively. For colony formation assays, 5 x 10^5 cells were overlaid with a mixture consisting of 5 μg pMEX*drm* or expression vector alone and 30 μl DOTAP (Boehringer Mannheim). After 6 hours this mixture was replaced with regular media and the cultures maintained for another 48 hours. Cells were then split 1:3, grown in the presence of G418 (Life Technologies; effective concentration, 400 μg/ml) for 2 weeks and colonies resistant to G418 were counted and isolated. Growth temperatures for transfected cells were: for REF-1 and CHO, 37°C; for DTM, 34°C; and for ST33c, 34°C and 39°C. Transient transfections of Cos-7 cells were performed using the pSVL vector expressing a HA-tagged *drm* and LipofectAMINE (Life Technologies, Gaithersburg, MD), according to the manufacturer's specifications.

***In situ* hybridization**. Tissues from Sprague-Dawley rats were processed and analyzed by *in situ* hybridization according to D. Sassoon (37). A non-radioactive riboprobe containing 1.9 kb of the 3' end of *drm* was generated by using Digoxigenin RNA Labeling Kit (SP6/T7) from Boehringer Mannheim, and concentration of the

labeled probe was determined by using the SIG Nucleic Acid Detection Kit (Boehringer Mannheim). Detection was performed by using Anti-Digoxigenin antibody, conjugated with Alkaline Phosphatase (Nucleic Acid Detection Kit, Boehringer Mannheim). Sections were counterstained with Methyl Green (1%) and mounted in Aqueous Mounting Medium (Signet Laboratories). Analysis was performed on a Nikon Labophot 2 microscope.

Analysis of apoptosis. ST33c cells were transfected with the control vector or with the vector containing *drm* at 34°C, and pools of G418-resistant colonies were selected, expanded and analyzed for expression of *drm*-specific mRNA. ST33c cells expressing *drm* were shifted to 39°C for 24 hrs, and cells were fixed in 3.7% formaldehyde in PBS (10 min, RT), washed three times, stained in DAPI (10 min, RT) and examined with a Nikon inverted microscope under UV illumination. DNA fragmentation analysis was performed as previously described (1).

Nucleotide sequence accession number. The *drm* sequence for the rat homologue has been assigned GenBank/EMBL accession number Y10019.

The characterization of a flat (non-transformed) revertant cell line, F-1, which was isolated from rat fibroblasts (DTM) transformed by the serine/threonine kinase oncogene *mos* has been previously reported (41). F-1 cells express high levels of v-*mos*-specific RNA and kinase activity, but fail to express characteristic transformed properties, including colony formation in soft agar and tumor formation in nude mice. Moreover, the revertants are resistant to re-transformation by v-*mos* and v-*raf*, while they can be efficiently transformed by v-*ras* and, with a somewhat lower efficiency, v-*src*. The reversion and resistance to re-transformation correlated with the failure of the serine/threonine kinase oncogenes v-*mos* and v-*raf* to activate the MAP kinase pathway due to their inability to activate MEK-1 or MEK-2, the immediate upstream activators of MAP kinase.

Since levels of MEK and MAP kinase were not changed in the revertant cells, and since growth factors and *ras* activated MEK and the MAP kinase cascade normally,

these results suggested that the reversion could be the result of mutations affecting the expression or function of genes which contribute to the activation of MEK by v-*mos* or v-*raf*, or from the expression in the revertant cells of genes which block this activation and which are down-regulated in DTM and other transformed cells. In an attempt to identify such transcriptional changes, differential display analysis was used to compare the expression of RNA in transformed and revertant cells. Described herein is the identification and characterization of a novel cDNA, designated *drm* (down-regulated in v-mos-transformed cells), which is expressed in the F-1 revertant and normal parental rat fibroblasts, but which is down-regulated in rat fibroblasts transformed by several retroviral oncogenes. The *drm* cDNA shows no significant homologies to known genes in DNA databases and contains an open reading frame (ORF) capable of encoding an 184 amino acid, cysteine-rich protein with a calculated molecular weight of 20,682. Regions of the drm protein show significant sequence homologies with the rat and human DAN (NO3) gene products (10, 28-30), which have been shown to possess tumor and growth-suppressing activities. The *drm* gene encodes a 20.7 kDa protein recognized by a specific antiserum in phenotypically normal rat cells. This protein was not detected in v-*mos*-transformed cells. Analysis of RNA from multiple tissues of the rat and *in situ* hybridization experiments in adult rats, indicate that *drm* expression is regulated in a tissue-specific manner. *In situ* analysis also indicate that *drm* RNA is predominantly expressed in terminally-differentiated, non-dividing cells, such as neurons, type-1 cells of the lung, and goblet cells of the intestine.

Transfection analysis demonstrates that *drm* overexpression in normal rat fibroblasts blocks cell proliferation, while co-transfection with *ras* oncogene reverses this inhibition. Furthermore, cells overexpressing *drm* and conditionally transformed with v-*mos*-expressing Moloney murine sarcoma virus (Mo-MuSV) rapidly undergo apoptosis when shifted to the non-permissive temperature. These results indicate that *drm* represents a newly identified gene which appears to play a role in cell growth and tissue-specific differentiation.

Identification of an mRNA expressed in revertant cells but repressed in v-*mos*-transformed rat fibroblasts. To identify genes expressed in F-1 revertant cells,

but not in v-*mos*-transformed parental cells (DTM), differential display analysis (25) was performed, using oligo dT-selected RNA isolated from rapidly-growing DTM and F-1 cells. Eight cDNAs showing differential intensities between DTM and F-1 mRNAs were identified and used to probe Northern blots containing poly(A)+ RNA from DTM and F-1 cells. Only one exhibited differential mRNA expression, detecting a 4.4 kb RNA expressed in F-1 cells, but absent in DTM cells. Analysis of this cDNA, designated *drm* (for down-regulated in v-mos transformed cells), revealed a 691 bp sequence, which included a consensus polyadenylation signal (AATAAA) located 20 bp upstream from the poly(A) tail, as well as the 5' and 3' primers used for PCR. A search of nucleotide sequences compiled in the GenBank data base showed no significant similarities to known genes.

Repression of *drm* mRNA expression following cell transformation. To establish a correlation between repression of *drm* gene expression and the transformed cell phenotype, the hybridization of *drm* cDNA to RNA from normal and transformed REF-1 cells was analyzed. *Drm* was expressed at similar levels in both REF-1 and revertant F-1 cells, but its expression was completely repressed in REF-1 cells transformed by the v-*ras*, v-*raf*, v-*src* and v-*fos* oncogenes. These results demonstrated that repression of *drm* expression was not restricted to transformation induced by v-*mos*.

Because the initial identification of *drm* was based on its expression in the F-1 revertant and it had been previously shown that F-1 cells could be transformed by v-*ras* and v-*src*, the effect of expression of these oncogenes in F-1 cells on *drm* expression was analyzed. F-1 cells expressing and transformed by v-*ras* and v-*src* did not contain *drm* transcripts detectable by Northern blot analysis, while in contrast, F-1 cells infected with the v-*mos* expressing MSV-124 show levels of *drm* RNA essentially identical to uninfected F-1 cells or REF-1 parental cells. Since it had been previously shown that superinfection of F-1 cells with additional copies of v-*mos* did not induce transformation (41), these results are consistent with the hypothesis that *drm* expression is down-regulated following oncogene-mediated transformation.

To further analyze the correlation between *drm* expression and the transformed phenotype, REF-1 cells transformed by a temperature-sensitive (*ts*) isolate of Moloney murine sarcoma virus (Mo-MuSV ts110) (3) were used. These cells (ST33c) are transformed at 34°C, but express a phenotypically normal, non-transformed phenotype at 39°C (42). Analysis of RNA extracted from cells maintained at both temperatures indicated that *drm* RNA was synthesized at 39°C in the absence of the v-*mos* protein and was markedly decreased at 34°C. Taken together, these results further indicate that in REF-1 cells repression of the *drm* RNA expression correlates with the transformed phenotype. The results with *ts* MuSV-transformed cells and the F-1 revertant indicate that *drm* expression is directly or indirectly modulated by the v-*mos* oncoprotein and its transforming functions.

***Drm* is a novel gene.** To fully characterize the *drm* gene and its product, rat fibroblast and rat kidney cDNA libraries were screened and five independent overlapping cDNA clones were isolated, which covered ~3820 bp of *drm* mRNA. Southern blot analysis indicated that the *drm* sequence is derived from a single gene spanning at least 12 kb and is not rearranged in either DTM, which does not express *drm*, or in the F1 revertant.

The 3820 nucleotides of cloned cDNA is shorter than the apparent size of the RNA identified in REF-1 cells, suggesting that the isolated clones may not include the entire *drm* mRNA sequence. However, this cDNA does contain a single long open reading frame (ORF) beginning at nucleotide 130 and terminating with an in-frame stop codon at nucleotide 693. Translation is predicted to start at the first in-frame methionine at nucleotide 139 within a favorable translation initiation context (A at -3, C at -4, G at -6 and A at +4) (22,23). Thus, the characterized *drm* cDNA consists of 138 bp of 5' untranslated (UTR) sequence (65% GC), a 552 bp coding region and 3130 bp of 3' UTR containing a consensus polyadenylation signal AATAAA located 21 nucleotides upstream from the poly(A) tail.

The major ORF contained in the *drm* cDNA would be predicted to encode a 184 amino-acid polypeptide with a calculated molecular weight of 20,682. The

presumptive *drm* gene product is highly basic (7.61% arginine, 8.7% lysine and 2.17% histidine), with the NH_2-terminal half containing a leucine-rich hydrophobic domain located between amino acids 4 and 24, whereas the carboxy-terminal moiety is characterized by the presence of nine cysteines. The presence of an amino-terminal hydrophobic domain suggested a possible membrane localization of the protein and analysis of the drm deduced amino-acid sequence using the TMbase database of transmembrane proteins (Lausanne) indicated a high probability that this protein could form a transmembrane helix in this region. Examination of the predicted sequence also identified two potential nuclear localization signals which fulfill the motif K(R/K) x (R/K): KPKK (amino acids 145-148) and KKKR (amino acids 166-169), two protein kinase C phosphorylation sites (TER, amino acids 84-86 and TKK, amino acids 165-167) and three cAMP and cGMP-dependent protein kinase phosphorylation sites (KKGS, amino acids 26-29, KKFT, amino acids 147-150 and KRVT, amino acids 168-171).

Comparison of the drm amino-acid sequence to the GenBank and EMBL data bases using FASTA program, showed that the drm protein exhibits an overall similarity of 30% with the rat and human DAN gene product, which expresses tumor-suppressive properties (28,29). Using the BLAST program, a 52% similarity was detected between the carboxy-terminal cysteine-rich half of drm, the central region of the DAN protein and the carboxy-terminal region of the Xenopus protein Cerberus (CER), a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer (4). Further analysis also revealed similarity to the carboxy-terminal cysteine-rich end of the human MUC2 intestinal mucin (16). The nine cysteines of the drm are also present in DAN, CER, and MUC2 gene products at similar amino-acid intervals. This alignment generated the cysteine motif CX13CX(8-9)CX3CX(14-18)CX2CX13CX(15-18)CXC. Within this motif several amino acids are conserved, suggesting that proteins containing this domain could be members of a related family.

Characterization of the *drm* gene product. *In vitro* transcription/translation of the ORF-containing 2.1 kb EcoRI fragment and 730 bp BamHI/KpnI fragment of *drm* cDNA confirmed that the presumptive open reading frame could express a protein

of approximately the expected size. To further characterize the *drm* product, an anti-peptide polyclonal rabbit antibody directed against amino acids 79 to 92 of the rat drm protein was generated. In order to assess the specificity of the antisera, an expression vector was constructed, synthesizing an epitope-tagged drm protein by introducing a DNA fragment encoding the nine-residue epitope of influenza virus hemagglutinin HA1 at the 3' end of the coding region. The pSVL expression vector containing this fusion was used to transfect Cos-7 cells and cell lysates were prepared 48 hrs later, immunoblotted on nitrocellulose filter and incubated with the drm antisera. A band with a predicted molecular weight of ~21.4 kDa was detected and the same band was revealed with the monoclonal antibody against HA tag. It was not detected when lysates were exposed to 990 antisera preincubated with peptide against which this antiserum was raised nor in lysates of cells transfected with an empty vector. A protein of the same molecular weight was detected in HA-*drm*-transfected Cos-7 lysates immunoprecipitated with 990 antiserum and blotted with anti-HA sera and this precipitation could be blocked by the homologous 990 peptide.

To identify the endogenous drm protein, total lysates from various cells were analyzed by Western blotting. Low levels of a 20.7 kDa protein were detected in primary embryonic rat fibroblasts and in REF-1 cells. Analysis of drm protein expression in ST33 cells, conditionally transformed by v-*mos*, showed good correlation with *drm*-specific RNA expression. The protein was not detected in lysates of transformed cells at 34°C, but could be seen in cell lysates prepared 48 hrs after shifting the cultures to the non-permissive temperature. Drm protein was not detected in lysates of v-*mos*-transformed DTM cells.

***Drm* RNA is expressed in a tissue-specific fashion in adult rats.** To further characterize the *drm* gene and its possible function, the expression pattern of *drm* was examined in rodent tissues. Northern blot analysis of polyA+ RNA extracted from adult rat tissues (Sprague-Dawley) showed that the *drm* gene was expressed in brain, kidney, spleen, testis and lung and was not detected in heart and skeletal muscle. Highest levels were seen in kidney, testis, brain and spleen, while levels in the liver and lung were significantly lower.

To investigate whether *drm* expression was specific for any particular cell type, tissues from the same strain of rat were analyzed by *in situ* hybridization using sense and antisense *drm* riboprobes. *In situ* expression patterns in general correlated well with the Northern analysis, but *drm* RNA appeared to be predominantly expressed in differentiated cells (e.g., neurons in brain, type 1 cells in lung, goblet cells in intestine). In all cases the control sense probe showed no detectable hybridization.

The brain exhibited ubiquitous expression of *drm* RNA. High levels of *drm* expression were found in both neurons and glial cells of the brain cortex, while in the cerebellum, *drm* RNA was strongly expressed in all cells of molecular and granular layers. Its expression was significantly weaker in Purkinje cells.

In the kidney, *drm* RNA was found in epithelial cells of the proximal and distal tubules in the cortex, medullae and papillae. Very strong signals appeared to be localized in the nuclei of the epithelial cells.

In the small and large intestine, the *drm* gene was predominantly detected in goblet cells and specifically in the most differentiated goblet cells (on the tip of the villi in small intestine and the base and neck of the crypt in large intestine). However, some goblet cells in the crypt of the small intestine were also found positive for *drm* expression.

In the lung, the *drm* expression was localized to the nucleus of type 1 epithelial cells lining the alveoli. Type 1 cells are known to be terminally differentiated from their precursor type 2 cells (6). *Drm* was not expressed in every type 1 cell, which could indicate a possible correlation of *drm* expression with the stage of cell differentiation. A few endothelial cells of the airways and a number of macrophages also expressed *drm* RNA, while in the spleen, *drm* RNA was detected only in megakaryocytes and in agreement with the results of Northern blot analysis, *drm* hybridization was not detected in liver, heart and skeletal muscle.

***Drm* blocks colony formation by normal, but not transformed cells.** To determine the biological effect of *drm* overexpression *in vivo*, a portion of the *drm* cDNA containing the full-length ORF was inserted into the *neo*-containing pMEX expression vector (21). This construct, as well as the empty vector, was introduced into REF-1 and DTM cells and G418-resistant colonies were counted after 2-3 weeks. Colony formation was inhibited 30-fold when REF-1 cells were transfected with the *drm* expression vector. The *mos*-transformed DTM cell colony formation was not affected. Similar results were also seen in CHO cells, indicating that inhibition of colony formation is not specific to REF-1 cells. Analysis of independent, *drm*-transfected G418-resistant clones of REF-1 cells showed that all surviving clones expressed very low or undetectable levels of exogenous *drm* mRNA, suggesting that survival may select for cells expressing low levels of *drm*. In contrast, DTM cells, which showed no inhibition of colony formation, exhibited high levels of exogenous *drm* expression. In some cases, expression of endogenous *drm* RNA was also increased in DTM cells expressing exogenous *drm*, suggesting a possible autoregulation loop of *drm* expression.

Since oncogene-transformed stable cell lines had shown down-regulation of *drm* expression (see above), the interactions between transforming oncogenes and *drm* were further investigated by co-transfecting REF-1 and CHO cells with *drm* and the activated (38) *ras* oncogene. Consistent with previous results with DTM cells, co-transfection of *drm* with the *ras* oncogene did not suppress morphological transformation. However, co-transfection of *ras* with *drm* reversed the *drm*-dependent inhibition of colony formation both in REF-1 cells (84% of the control) and in CHO cells. The level of exogenous *drm* RNA in 5 of 6 G418-resistant clones co-transfected with pMEX*drm* and *ras* was increased. These data are consistent with the hypothesis that high levels of *drm* inhibit the growth or viability of normal cells, but that transformed cells are resistant to this inhibitory effect.

Conditionally-transformed cells expressing exogenous *drm* undergo apoptosis at the non-permissive temperature. Since transfection of non-transformed rat and hamster cells with *drm* expression vectors leads to the inhibition of cell growth,

stable cell lines expressing high levels of *drm* could not be obtained for molecular and biological analysis. In order to overcome this problem, conditionally-transformed ST33c cells were used to investigate the effects of *drm* overexpression. When v-*mos* is functional (34°C) and ST33c cells are transformed, transfection of pMEX*drm* vector does not affect the efficiency of colony formation in comparison to control vector. These results are consistent with the data for DTM cells and for REF-1 cells co-transfected with pMEX*drm* and *ras*, showing that the presence of transforming oncogene blocks the inhibitory effect of *drm*. In contrast, at 39°C, the percentage of survived colonies following pMEX*drm* transfection was significantly lower than that observed in control vector-transfected ST33c cells.

To analyze how *drm* overexpression blocks cell growth and colony formation, G418-resistant colonies of transfected ST33c cells were isolated at 34°C and tested for the expression of *drm*. Pools of G418-resistant cells expressed elevated levels of *drm* RNA similar to those seen in transfected DTM or *ras*-transformed cells. These transfected pools grew like the parental ST33c cells at 34°C, when v-*mos* is expressed, but rapidly lost viability after shifting to 39°C, and colony-forming ability was significantly reduced. This is consistent with the fact that, as previously shown, v-*mos* is not expressed in these cells at 39°C, and thus cannot neutralize the effects of the high level of exogenous *drm* in these cells. The morphological changes seen in these cells at 39°C resemble those of cells undergoing apoptosis, including cell shrinkage, cell membrane blebbing and loss of cell-cell contact and adhesion to the substrate. Furthermore, *drm*-expressing ST33c cells exhibited nuclear fragmentation and condensation within 24 hrs of a shift to 39°C, while no such fragmented nuclei were observed in these cells cultured at 34°C or in REF-1 cells at either 34° or 39°C. It was observed that 15-30% of the ST33c cells expressing *drm* at 39°C exhibited fragmented, condensed nuclei, while only 5-6% of the control ST33c cells manifested similar changes following a shift to 39°C. DTM cells, transfected with *drm* and containing two copies of v-*mos* (ts- and w.t. v-*mos*) also showed 5-7% fragmented nuclei at 39°C, which could represent the background level for *ts* v-mos-transformed cells shifted to 39°C. Apoptosis of *drm*-expressing ST33c cells at 39°C was also confirmed by agarose gel electrophoresis of genomic DNA, which showed significant fragmentation

only in the cells shifted to 39°C. Furthermore, the relative fraction of cells undergoing apoptosis were seen to correlate with the level of *drm* expression in a series of individual clones of ST33c cells transfected with *drm*. Taken together, these data suggest that cells expressing high levels of *drm* undergo apoptotic death in the absence of oncogene-induced transformation.

Example II. Isolation and characterization of human *drm* gene and gene product.

Cell culture, transfection and synchronization. All human cells, including normal diploid fibroblasts, were grown in HG-DMEM. CHO cells were grown in F12 medium. All media was supplemented with 10% fetal calf serum (FCS) (Atlanta Biological, Norcross, GA) and cells were maintained at 37°C with 10% or 5% CO_2 (for CHO cells). For serum starvation, medium was changed to 0.1% FCS when cells were subconfluent and cells were left in this medium for 72 hours. For density-dependent inhibition, cells were plated at $10^4/cm^2$ in 10% FCS. Twenty-four hours after plating, the medium was changed every two days. Exponentially-growing cells are cells cultured for 24 hours in 10% FCS. Human cells were synchronized as described previously (71). Briefly, IMR90 or Hem cells were grown in MEM α modification (Gibco, BRL) with 0.1% FCS for 72 hours prior to replacement with 10% FCS. Nine hours later, hydroxyurea (HU) (Sigma) was added to a final concentration of 0.5 mmol/U to arrest the cells at the G_1/S boundary. After nine hours of HU blockade, the complete medium was added and cells were taken for protein and flow cytometry analysis (FACS).

Transient transfections of cells were performed by using Lipofect AMINE or Lipofect AMINE PLUS (for IMR90) (Life Technologies) as specified by the manufacturer.

FACS. For cell cycle analysis of human cells, at hourly intervals, the cells were harvested and washed with PBS, the number of cells was counted and 1 x 10^6 cells were processed for flow cytometry. Cells were suspended in PBS with 0.05% Triton X100. DNase-free RNase (200 U/ml, Boehringer Mannheim) was added for 30

minutes at 37°C and then the cells were washed twice. Propidium Iodide (PI) was added to a final concentration of 50 mg/ml (71). The cells were examined for DNA content with FACScan flow cytometer (Coulter Epic S' Profile II, Coulter Corp., Miami, FL) and the percentages of cells in G_0/G_1, S and G_2/M phases were determined with MultiPlus AV version 3.0 software.

To analyze the cell cycle of sorted cells, CHO cells were transfected with pEGFP or pDRM-GFP. At 24 hours after transfection, cells (50×10^6) were harvested by trypsinization and EGFP-expressing cells were recovered by fluorescence-activated cell sorting (FACS). Cells were fixed in 70% ethanol at 4°C and recovered by centrifugation. The fixed cell pellet was resuspended in 0.9 ml of PBS with 0.1% BSA and RNaseIIIA (200 U/ml) was added for 15 minutes at RT. DNA was stained with PI and examined with FACScan flow cytometer (Coulter Epics 753, Coulter Corp., Miami, FL), and the percentages of cells in G_0/G_1 S and G_2/M phases were determined with MultiPlus AV, version 3.0 and Elite software programs.

Northern blot analysis. For Northern blot analysis, Human Multiple Tissue Northern (MTN) blots (I-II), (II-III) (Clontech) and human RNA master blots (Clontech) were used. The blots were probed with a radiolabeled human DRM-specific probe. Hybridization and washing conditions were in accordance with the manufacturer's instructions.

Total RNA was extracted from cultured cells by RNAzol B (Tel-Test, Inc., Friendswood, TX), and hybridized with a human DRM probe as described previously (Topol *et al.*, 1992).

Screening of a cDNA library. To determine the DRM cDNA sequence, a human small intestine 5'-stretch cDNA library in λgt11 (Clontech) was screened using 5' sequences of rat *drm* (Cl.7ZAP) (65). Five clones were isolated. The largest one (3.2 kb) was amplified and analyzed. Both strands of the double-stranded plasmid DNA were sequenced by primer walking using the dideoxy chain dye terminator method with Amplitaq DNA polymerase, FS (Perkin Elmer). The sequencing products

were analyzed on an ABI prism 377 DNA sequencer (Perkin Elmer). The nucleic acid sequence of the DRM gene was analyzed using the GCG package (University of Wisconsin).

Rapid amplification of cDNA ends (RACE). For 5'-RACE, 1 μg of total RNA from human diploid fibroblasts was mixed with the DRM-specific primer and reverse transcribed with 200 U of Superscript II reverse transcriptase (Gibco/BRL) at 42°C for 30 minutes according to the manufacturer's protocol. The final products were subcloned into the EcoRI site of the pCRII plasmid and sequenced with vector-specific oligonucleotide primers.

Construction of EGFP-DRM fusion expression vector. The coding region of the DRM gene was PCR amplified from a cDNA using Ultima DNA polymerase (Cetus) and primers containing a BamHI restriction site. The primers used were 5' (CGGGATCCAGAATGAATCGCACGGCATAC) (SEQ ID NO:11) and 3' (GCGGATCCTTAATCCAAGTCGATGGATATGC) (SEQ ID NO:12) (primers from Biosynthesis, Inc., Lewisville, TX). The PCR product was digested with BamHI and inserted into an EGFP-C1 expression vector (Clontech) which was digested with BamHI and treated with Shrimp Alkaline Phosphatase (Boehringer Mannheim).

Western blot analysis. Cells were lysed in boiling 2x SDS sample buffer. Equal amounts of lysates (determined by Bradford protein staining reagent, Bio-Rad) were electrophoresed on 4-20% SDS-PAGE and transferred to Hybond ECL nitrocellulose membrane (Amersham). Equal loading and transfer was confirmed by staining reversibly in 0.2% Ponceau - 6% TCA (Sigma). The membranes were incubated first with the appropriate primary antibody and then with horseradish peroxidase-labeled secondary antibodies (Amersham). Antibodies were detected by using the ECL detection system (Amersham) or the Super Signal CL-HRP Substrate System (Pierce) and visualized by using Kodak XAR-5 X-ray film. Western blots were stripped for reprobing with other primary antibodies as specified by the manufacturer (Amersham).

Probes and antibodies. cDNA probes were obtained from the following sources: rat NSE cDNA (79) from Dr. Gregor Sutcliffe; human GFAP cDNA was purchased from the ATCC. Polyclonal antibodies (e.g., 990), which recognized DRM, were described previously (65). Other antibodies used in this study were specific for $p27^{Kip1}$, $p21^{Waf1}$, cyclin E (Transduction Lab., Lexington, KY), cyclin E (Ab-1, Oncogene Research), cyclin E (M-20, Santa Cruz Biotechnology; SC35), cyclin D1 (R-124, Santa Cruz), GFP (Clontech), p53 (PAb122, D01, Pharmingen), pCdK2 (M2, Santa Cruz), PhosphoPlus Rb/Ser 795), antibody kit (New England Biolabs), β-actin (Chemicon).

BrdU incorporation. The effect of DRM expression on bromodeoxyuridine (BrdU) incorporation was determined in CHO cells growing asynchronously in F-12 - 10% FCS. Cells were plated at 10,000 cells/ml on coverslips and after 24 hours were transfected with 5 μg of either pEGFP, or pDRM-EGFP. Twenty-four hours after transfection, the medium was changed and cells were incubated with BrdU labeling reagent for a further 12 hours according to the supplier's (Amersham) instructions. After labeling, coverslips were washed in PBS and cells were fixed in 3% paraformaldehyde. Incorporated BrdU was detected with a monoclonal anti-BrdU antibody (Boehringer Mannheim) by immunocytochemistry.

Immunocytochemistry and immunofluorescence. Fixed cells on coverslips were washed twice with PBS and treated with 0.1M glycine in PBS for 5 minutes at RT, followed by treatment with 0.1% Triton X-100 in PBS for 4 minutes at RT and 50 mM NaOH for 10 seconds. Co-localization of DRM with the speckles was analyzed by immunofluorescence with a monoclonal antibody SC35 (80) and a rhodamine-conjugated, goat anti-mouse immunoglobulin G secondary antibody (Kirkegaard and Perry Labs., Gaithersburg, MD). Coverslips were mounted and examined with a fluorescence microscope.

Chromosomal mapping of DRM gene. A somatic cell hybrid panel (Oncor) was hybridized with a ^{32}P-labeled 1.2 kb human 5' DRM cDNA fragment according to the manufacturer's protocol.

In order to localize the DRM gene on human chromosomes, a special probe was prepared by PCR using primer #197 (position 2934-2955): 5'TCATTACATCATCAGTGACTCG3' (SEQ ID NO: 20) and #195 (position 3131-3152): 5'CAGATTTGGCTCAAGTAAAGAG3' (SEQ ID NO:21). The result of this reaction was a fragment (195 PCR) representing 218 bp specific for the human DRM sequence. Chromosomal localization of the 195 PCR product was accomplished using two panels of somatic cell hybrids. The first was a hybrid mapping panel #2 from the Coriell Institute for Medical Research. This is a collection of 24 human X hamster cell lines. All but two of these hybrids retain a single, intact human chromosome. The second panel is the GenBridge 4 radiation hybrid panel available from Research Genetics (73). PCR reactions were carried out as follows. Twenty-five ngm of hybrid or control DNA were amplified in a 10 μl volume in a reaction buffer consisting of 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM of each dNTP, 1 pmol of each primer and 0.001 units of Taq Gold (Perkin Elmer) polymerase. The PCR cycling conditions were as follows: an initial 94°C denaturation step for 10 min followed by 35 cycles of 94°C denaturation for 30 sec, 60°C annealing for 1 min and a 72°C extension step for 1 min, followed by a 72°C heating for 5 min. PCR products were run out in 1.2% agarose gels and stained with ethidium bromide. After scoring each radiation hybrid for the presence or absence of the PCR product, the resulting vector was sent by electronic mail to the MIT/Whitehead Institute Genome Center for analysis.

Subcellular Fractionation. Subcellular fractionations were prepared as described previously (89). The fractionation protocol was first verified on COS7 cells transfected with expressing vector pGFP (Green Fluorescent Protein) to confirm the correct distribution of control proteins. Cells grown on 100 mm culture dishes as a monolayer were washed and scraped in PBS, centrifuged and resuspended in hypotonic buffer A (10 mM Hepes, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM PMSF) (18). After 15 min of swelling on ice, cells were homogenized carefully by 20-25 strokes in a Dounce homogenizer (Type B pestil) to break the cells. This procedure was carefully monitored by fluorescence microscopy for staining of "broken cells" with propidium iodate (PI) to ensure >90% lysis of the cells without breakage of the nuclei. After centrifugation at 800g for 10 min (4 C), the pellet, consisting of a mixture of unbroken

cells and crude nuclei, was designated the low speed pellet and was processed further. The supernatant was collected and subjected to further centrifugation at 100,000g for 30 min. The resulting supernatant contained soluble protein and was designated the cytoplasm fraction (C). The pellet was considered the particular fraction (P). The low speed pellet was washed in a large volume of buffer A and resuspended in 2 vol buffer A' (buffer A supplemented with 0.5 mM DTT and 1% NP-40) of the initial cell pellet. After incubation on ice for 10 min, the sample was centrifuged, the supernatant was removed and cleared as described above, generating a pellet (N) and supernatant fraction. This resulting supernatant, containing soluble cytoskeleton proteins, was designated the skeleton fraction (Sk). The pellet (Pk) represented unsoluble cytoskeleton fraction. The remaining nuclei were again washed in Buffer A', pelleted at 10,000g, resuspended in 4 vol 2xSDS-loading buffer, sonicated three times for 20 s, and boiled for 10 min. Each subcellular fraction was then assayed for its protein content and an equal amount of total protein (40 g) was loaded on the gel.

Molecular cloning of human DRM. A new gene sequence (*drm*) (GenBank Accession No. Y10019) has been previously identified, based on differential display analysis of v-*mos*-transformed rat fibroblasts and their flat revertant (65). Zoo-blot analysis indicated that the *drm* sequence is present not only in rodents (rat and mouse) but also in humans. To isolate the human *drm* homolog a human small intestine 5'-stretch cDNA library was screened with a probe that encompasses the coding region of rat *drm* to obtain a full-length of cDNA insert. Among the positives, the longest clone (3.2 kb) found included the majority of the open reading frame (ORF) of *drm*. To extend the 5' end of the obtained clone the 5' RACE-PCR technique was applied on RNA extracted from primary human diploid fibroblasts and extended the clone for an additional 200 bp. This 3.411-nucleotide sequence, excluding the poly(A) tail, contains one large ORF from position 130 to 683, which encodes a protein of 184 amino acids (M_r, 20, 682). A single ORF was found, with the ATG translation initiation site located at position +1 and the TAA stop codon at position +553. This ORF is preceded by a stop codon (TAG) at position -105. This was designated as the translation start site as there was no ORF upstream of this codon and it includes a Kozak consensus sequence for translation initiation (74).

Comparison of the human and rat DRM cDNAs revealed that these two cDNAs have a highly-related sequence in the coding region (~86% identity), but they are divergent in 5' and 3' untranslated sequences (UTR). In the 5' UTR, the hu-DRM contains two long stretches of GC (19 and 11 nucleotides) at -100 and -80, respectively. Comparison of the rat and human DRM amino acid sequences demonstrated a high conservation (181/184 amino acids) between rodent and man. Like rat drm, human DRM has two putative nuclear localization signals near the C-terminus (amino acids 145-148 and 166-169), a cysteine-rich region (93-178) and several sites for phosphorylation by protein kinase C (amino acids 84-86, 165-167), cyclic AMP and cyclic GMP-dependent protein kinases (amino acids 26-29, 147-150 and 168-171), respectively. This striking identity implies that the overall three-dimensional shapes of the two proteins are very similar. This may in turn indicate that the two proteins are functionally equivalent.

DRM maps to human chromosome 15. Southern blot analysis of BamHI-digested DNA from mouse-human somatic cell hybrids harboring a single human chromosome was carried out using 1.2 kb human DRM 5' cDNA as a probe. One single band was detected in the DNA from hybrid cells harboring human chromosome 15. The DRM gene was also localized by PCR analysis.

Successful amplification of the 218 bp human 195 PCR product was obtained in control human, but not in hamster DNA. Amplification of the Cornell hybrid DNA indicated that this gene was located on chromosome 15. Analysis of the radiation hybrid data placed this PCR product 23.32 cR distal to the chromosome 15 reference marker WI-5590 and one cR distal to marker D15S144. This is a position about 59 cR from the top of the chromosome 15 radiation hybrid map, about 23 cM from the top of the linkage map and corresponds to a cytogenetic location of 15q11-q13 (73,75).

DRM is a secreted protein that remains cell associated. The cellular localization of DRM has also been analyzed using both cell fractionation and immunofluorescence microscopy. COS cells transfected with pHA-DRM were separated into multiple subcellular fractions and the relative distribution in the

particulate (P), soluble cytoplasmic (C), nucleus/cytoskeleton-associated soluble (Sk) and insoluble (Pk), and pure nuclear (N) fractions, was determined by western blot analysis with anti-DRM antibodies. The protein was detected predominantly in the insoluble particulate fraction (P) and the detergent-extracted soluble and insoluble cytoskeleton-associated fractions (Sk and Pk). Quantitation of these results by densitometry indicated that over 70% of DRM was localized in the insoluble membrane and cytoskeletal fractions (Pk and Sk), while 17% was found in the cytoplasmic (C) fraction and 9% in the nucleus (N). To verify the subcellular fractionation, the same filters were blotted with antibodies recognizing the membrane localized p145 c-met protein. As expected, c-met was found predominantly in the insoluble membrane fraction (fraction P).

To confirm and further analyze the distribution of DRM, DRM localization in COS cells overexpressing pHA-DRM was investigated by immunofluorescence. Transfected cells were fixed with paraformaldehyde and probed with DRM polyclonal antibodies and Oregon green 488 conjugated anti-rabbit secondary antibody. Alternatively, the cells were permeabilized following fixation and subsequently treated with antibodies. Permeabilized cells exhibited a diffuse, fiber-like network of staining, suggestive of a localization in the endoplasmic reticulum/Golgi complex, and some cells also exhibited a distinct perinuclear staining, which could be the site of DRM synthesis. To confirm this intracellular localization, monoclonal antibodies directed against the Golgi-specific p58K protein, specifically localized on the cis/medial side of the Golgi apparatus were used. The results showed that both DRM and p58K co-localized in the Golgi stacks.

In contrast, non-permeabilized cells showed a clumped, punctate pattern that appeared to surround the outer surface of the cell membrane, indicating the presence of DRM on the external cell surface. Analysis of live, unfixed cells showed a similar pattern. A similar subcellular distribution of DRM was observed in COS cells by using anti-HA antibodies and in rat cells expressing the endogenous protein, although in the latter, intracellular staining was predominantly cytoplasmic and perinuclear.

Taken together, these results indicate that DRM is transported through the cell membrane to the outer surface of the cell. To confirm that the hydrophobic region was responsible for DRM's entrance into the secretory pathway, COS7 cells were transfected with pHA-DRM-21N and the localization of the truncated protein was determned by using anti-DRM and anti-HA antibodies. The truncated protein was found to be exclusively intranuclear, consistent with the fact that the protein also contains 2 NLS's (amino acids 147-150 and 168-171), and indicating that the two NLS signals are functional. As expected, surface staining was not observed when these live or nonpermeabilized cells were treated with antibodies, indicating that DRM is unable to be secreted in the absence of the 21aa amino terminal region.

Results of both cell fractionation and immunofluorescence indicated that DRM is a secreted protein. However, the protein was not detected in culture fluids of either COS7 cells overexpressing DRM, CHO cells expressing transfected DRM, or rat fibroblasts expressing the endogenous protein. The failure to detect soluble DRM was not technical because the reconstitution experiments demonstrated that the protein was detectable under these conditions. To test the possibility that the secreted DRM protein remains associated with the external cell surface, pHA-DRM transfected COS cells were treated with acidic buffer, conditions which have been shown to dissociate non-covalently bound polypeptide ligands from their receptors. This treatment significantly reduced the amount of detectable glycosylated DRM, whereas it did not apparently decrease the amount of the faster migrating non-glycosylated form.

When transfected CHO cells were treated with acid buffer, the amount of DRM proteins significantly decreased and the upper glycosylated band was no longer detectable. Treatment of both transfected cell lines with trypsin decreased the amount of glycosylated DRM. Incubation of the same membranes with anti-EGF-R or actin antibodies showed that the levels of these two proteins were not affected by these treatments. To confirm that intact DRM protein had been removed from the outer plasma membrane, proteins were concentrated in the acid wash by acetone precipitation and analyzed by immunoblotting. The protein was detectable in the acetone-precipitated sample at low levels, migrating as multiple bands.

The DRM/GFP fusion protein is a nuclear protein. In order to localize the DRM product a vector containing the fusion EGFP-DRM insert under a CMV promoter was constructed. CHO cells were transfected with the expression vectors encoding only green fluorescent protein (pEGFP) or fusion EGFP-DRM (pEGFP-DRM). Comparison of the fluorescence from the EGFP alone with that of the EGFP-DRM fusion showed that the chimeric protein was exclusively localized in the nuclei of CHO cells. EGFP-DRM product was also found to be localized in the nuclei of HeLa, SaoS, Cos-7, and normal human fibroblasts transiently transfected with EGFP-DRM vector. The pattern of distribution of EGFP-DRM in the nuclei varies, including, predominantly, structures of punctate shape (dots), but very rarely, in single cells, uniformly diffused nuclear distribution could be seen. Amounts of nuclear dots could be different: from a few large to numerous small ones. Taking into account this specific pattern of distribution in the nuclei which resemble a speckled pattern, experiments were conducted to co-localize DRM with other known subnuclear structures such as non-snRNA splicing factors (SC35) (81). In immunofluorescence labeling experiments with monoclonal anti-SC35 antibody for transiently-transfected Cos cells with GFP-tagged DRM, SC35 and DRM did not co-localize, but in several nuclei these two proteins did occupy the same regions. DRM did not co-localize with nucleoli, as determined by co-transfection of HeLa and CHO cells with blue fluorescent protein (BFP)-tagged Rev, which is known to have nucleoli localization (82).

Distribution of DRM transcript in normal human tissues. To characterize the level of endogenous DRM mRNA expression in human tissues a multitissue poly(A)+ RNA Northern blot (Clontech) was hybridized with a 1.2 kb 5' end hu-DRM cDNA fragment. On a Northern blot, a single transcript of approximately 4.4 kb was detected in several tissues, including the prostate, ovary, small intestine, colon, brain, skeletal muscle and pancreas. The highest level was seen in the small intestine and colon; however, in the brain and ovary, DRM expression was also high based on normalization of poly(A)+ RNA for β-actin. No specific mRNA was detected in spleen, thymus, heart, lung, liver, placenta and peripheral blood leukocytes. This expression pattern of DRM is different from the expression pattern of the rat DRM, but in both, the brain was positive for DRM expression. To expand the information about

the tissues where DRM is expressed, the human RNA Master Blot was used, whose data confirmed the previous one, but showed that DRM also is expressed in colon, stomach, appendix and lymph nodes.

To investigate whether DRM expression could be detected during human embryonal development, a human fetal multiple tissue Northern blot (Clontech) was analyzed, demonstrating that DRM is highly expressed only in fetal brain. Previously, using *in situ* hybridization, it was shown that the rat adult brain exhibited ubiquitous expression of *drm* RNA (65). The expression of human DRM in different regions of the human brain was examined. The analysis of several human brain regions revealed widespread expression of DRM, although with different intensity. Based on normalization for β-actin, the highest abundance was found in the putamen, corpus callosum, substantia nigra, caudate nucleus and cerebral cortex. A high level of expression was found in the medulla, thalamus and subthalamic nucleus, and a low level of expression was detected in the amygdala, spinal cord and frontal lobe.

Based on previous data in rat (65) where a high level of DRM expression was detected in neurons, a specific marker for neurons, neuron-specific Enolase, NSE (79) and glial fibrillary acidic protein, GFAP (84), was used as a marker for astrocytes, to evaluate the connection of DRM expression with these two markers. In corpus collosum, the major expression of DRM-specific RNA coincides with a high level of GFAP expression, which is specific for astrocytes. At the same time, in the cerebellum and cerebral cortex, a high level of DRM expression coincides with expression for a neuron marker, which supports the data obtained with *in situ* hybridization earlier. In putamen, temporal lobe, frontal lobe and occipital pole, all DRM expression coincides with NSE, which suggests that DRM is expressed in differentiated neurons in the adult human brain.

DRM expression in normal and transformed cultured cell lines. Since DRM was initially isolated as a gene whose expression was down-regulated in v-*mos*-transformed cells, more than 70 human tumor and normal diploid cell lines were screened for DRM expression. The DRM transcript was found predominantly in

normal human diploid fibroblasts of different origins (10/10) and in normal human astrocytes, but was not detected in normal melanocytes, normal mammary glands and the HUVEC cell line. DRM was not detected in essentially all tumor cell lines examined. These results raised the possibility that the tumorigenic phenotype is incompatible with the continued expression of DRM and that down-regulation of DRM is necessary as a step in transformation. To investigate this assumption, the level of DRM expression in cells was examined at different stages of transformation. We established a system containing primary, immortalized and transformed rat fibroblasts, isolated RNAs and proteins from the cells and determined the level of DRM expression. Primary rat fibroblasts were shown to contain a high level of DRM on RNA and protein levels; in immortalized cells (REF-1) the level of DRM was decreased 2-fold. Finally, in transformed rat fibroblasts the DRM expression was not detected at either RNA and protein levels. These results demonstrate that the level of DRM expression is tightly regulated and may reflect both the state of transformation and/or proliferative activity.

To assess the expression of DRM during density-dependent growth inhibition, normal human fibroblasts were seeded in 10% FCS and the medium was replaced every second day with fresh 10% FCS. Northern blot analysis showed DRM induction after 6 days of density inhibition of growth when cells entered quiescence. Most striking is the fact that the expression of DRM-specific RNA was amplified up to 10-fold in density-arrested human fibroblasts. These data demonstrate that human fibroblasts accumulate DRM mRNA when they exit the cell cycle and enter a quiescent state as they grow to high density.

Modulation of DRM expression during the cell cycle. Since DRM expression was found to increase in primary rat fibroblasts when proliferation is under strong regulation and in human fibroblasts under density-mediated arrest in G_0, the DRM protein level was examined for changes during the cell cycle. Normal human diploid fibroblasts (IMR90 and HEM cells) were synchronized by serum starvation for 72 hours in minimum essential medium alpha modification (71) followed by arrest at the G_1/S boundary by hydroxyurea (HU) blockade and subsequent release of this block

with fresh complete medium. Lysates were prepared at different times after HU blockade release and samples were analyzed by Western blotting with anti-DRM antibodies. It appears that the level of DRM proteins change in a cell cycle-dependent manner. The highest amount of DRM was observed during G_0 when the cells were arrested by serum deprivation for 72 hours. The level of DRM protein was found to decrease 3-fold as cells reached the G_1/S boundary, to be low during the S phase and to increase again in the end of the S phase and as cells entered the G_2/M phase. Cyclin E expression was used as a control for cell cycle progression (78). The changes in DRM levels do not correlate with the changes in DRM in the RNA level. Fluorescence-activated cell sorting (FACS) analysis with parallel cultures, indicated that cells enter the S phase at 1 hour after HU blockade release under these experimental conditions. The experiment was repeated with HEM cells and the results were consistent with previous findings. These data indicate that the level of DRM declines when cells enter the S phase of cell cycle. In order to see the early response of DRM expression just after addition of a mitogen, HEM cells were growth arrested by serum starvation and reintroduced into a synchronous cell division cycle by addition of 10% FCS. By this method, it was shown that biosynthesis of DRM is clearly down-regulated 1.5 hours after serum stimulation.

Several proteins that are involved in the cell cycle regulation are accumulated during starvation such as $p27^{Kip1}$ (76) and cyclin E (86). The pattern of modulation of DRM during the cell cycle was compared with other inhibitors. Whereas p27 tends to accumulate in quiescent cells and declines in response to mitogenic stimulation, p21 levels are generally low in quiescent cells, but rise in response to mitogen treatment.

The pattern of DRM expression during the cell cycle and the first three hours of serum stimulation is very similar to that observed for $p27^{Kip1}$, but contrasts to $p21^{Cip1}$. Although the amount of DRM falls significantly during the G_0 to S phase transition, it continues to be synthesized in proliferating cells, leaving the possibility open that its expression might also be regulated periodically.

Previously, it was known that cell cycle regulation of many proteins, such as cyclins, cyclin-dependent kinase inhibitors, p27, occurs via the ubiquitin-proteosome pathway. Also, it has been shown that compared to proliferating cells, quiescent cells contain a far lower amount of p27 ubiquitinating activity (76,77). In order to test a hypothesis that accumulation of DRM in starved cells is also due to increased stability of the protein in quiescent cells, the effect of the proteosome inhibitors, lactocystin (LC) and clasto-lactocystin-β-lactone, and chloroquine, the lysosomal inhibitor was examined.

Degradation of DRM Proteins. To study the stability and maturation of DRM and monitor the appearance of DRM forms, pulse-chase experiments were performed in primary rat fibroblasts. Cells metabolically labeled with 35S cysteine for 30 min were either lysed immediately (pulse) or incubated in excess of cold cysteine for various periods of time (chase). DRM protein was immunoprecipitated with specific antiserum and immune complexes were separated on SDS-PAGE. Both glycosylated and non-glycosylated forms were detected after a 30 min pulse. The same bands were visible when the pulse period was shortened to 10 min, indicating that glycosylation takes place during or immediately after biosynthesis. Intensity of the labeled bands rapidly decreased over a two-hour chase period, in agreement with an estimated half-life of about 45-60 min. Both glycosylated and non-glycosylated forms were lost at equivalent rates, indicating that glycosylation did not influence protein stability. A mobility shift of all DRM bands was also observed that was visible after a 30 min chase, suggesting that phosphorylation is involved in degradation. To confirm that the shifted bands were indeed phosphorylated, cell extracts were treated after a 30 min pulse and after a 2.5 h chase period with alkaline phosphatase. All DRM bands were sensitive to this treatment, especially after the 2.5 h chase, as shown by their increased electrophoretic mobility.

To determine which of the endosomal/lysosomal or proteasome pathways was involved in DRM protein degradation, pulse chase experiments were performed in the presence of either chloroquine, a lysosomotrophic protein inhibitor or lactacystin, a specific inhibitor of proteasomal degradation. Protein stability was observed to be

increased in the presence of both inhibitors, although the observed relative intensity of the upper and lower bands, as well as their mobility, depended on the inhibitor used. Thus, in the presence of chloroquine, the stability of the glycosylated form was apparently increased, compared to that of untreated cells and of the lower non-glycosylated form. In addition, the mobility of the upper stabilized band was increased, suggesting it may have undergone dephosphorylation. These changes are consistent with the hypothesis that phosphatase activity in lysosomes acts to dephosphorylate DRM during treatment. In contrast, in the presence of lactacystin the stability of the lower non-glycosylated form was increased. Moreover, changes in mobility were not observed, suggesting that phosphorylation of all forms was preserved, possibly as a signal for degradation by proteasomes.

Example III. Production of EGFP/DRM fusion proteins

The EGFP/DRM fusion encoding nucleic acid (SEQ ID NO:1) was constructed as follows: DRM was PCR amplified using: forward primer: CGGGATCCAGAATGAATCGCACGGCATAC (SEQ ID NO:11) and reverse primer: GCGGATCCTTAATCCAAGTCGATGGATATGC (SEQ ID NO:12). The PCR product was digested with BamHI and EcoRI and ligated in frame into the pEGFP-C1 vector digested with BglII and EcoRI. The EGFPC1 coding region is nucleotides 3954-4688 and the DRM coding region is nucleotides 4689-5243. The amino acid sequence of the EGFP/DRM fusion protein is SEQ ID NO:29.

The NUCLEAR LOCALIZATION MUTANT #1(NLS#1), which contains a deletion of the 3' NLS region of DRM was made by cutting the EGFP/DRM fusion gene (SEQ ID NO:1) with BstXI and ligating in the double stranded synthetic oligonucleotide:

TAAGTCGCTTCGACGTACATTCAGCGA (SEQ ID NO:13)

to remove the 3' portion of the *drm* gene including the 3' nuclear localization signal (NLS#1) but leaving the 5' nuclear localization signal (NLS#2). The EGFP coding region is nucleotides 3954-4688 and the *drm* N1 mutation coding region is nucleotides 4689-5147. The resulting nucleic acid sequence is SEQ ID NO:5. The amino acid sequence of the NLS#1 mutant is SEQ ID NO:30.

The NUCLEAR LOCALIZATION MUTANT #2 (NLS#2), an EGFP-DRM double mutant, contains a deletion of the 3' NLS#1 and a point mutation within the upstream NLS#2. The EGFP coding region is nucleotides 613-1338 and the *drm* 2nls mutant coding region is nucleotides 1339-1815. This mutant was generated by PCR amplification of *drm* with the 5' oligonucleotide: AGGAATTCAATGAATCGCACGGCATAC (SEQ ID NO:14) and the 3' reverse oligonucleotide primer: ACGGGATCCTTACATGGTGGTGAATACTTGGG (SEQ ID NO:15), which introduces a point mutation in the 5' NLS#2, rendering it non-functional. The resulting nucleic acid sequence is SEQ ID NO:6 and the amino acid sequence of the NLS#2 mutant is SEQ ID NO:31. This PCR-generated fragment was digested with restriction enzymes BamHI and EcoRI and ligated into a BamHI and EcoRI digested EGFP-C1 vector obtained from Clontech Inc.

Generation of D5del versions of EGFP-DRM and NLS mutants:

D5del: The EGFP-DRM nucleotide sequence (SEQ ID NO:1) was digested with BsrGI and Bpu1102I. The double stranded synthetic oligonucleotide:

GTACAAGTCCGGACTCAGAATGAGGGCTTCAGGCCTGAGTCT TACTCCCGAGT (SEQ ID NO:16)

was ligated into the digested plasmid producing a EGFP-*drm* fusion minus the transmembrane domain. The EGFP coding region is nucleotides 3954-4682 and the *drm* coding region is nucleotides 4683-5129. The resulting nucleic acid is SEQ ID NO:7 and the amino acid sequence of the D5del mutant is SEQ ID NO:32.

NLS#1D5del: The EGFP-NLS#1 mutant nucleotide sequence (SEQ ID NO:5) was digested with BsrGI and Bpu1102I. The double stranded synthetic oligonucleotide:

GTACAAGTCCGGACTCAGAATGAGGGCTTCAGGCCTGAGTCT TACTCCCGAGT (SEQ ID NO:17)

was ligated into the digested plasmid producing a EGFP-*drm* fusion minus the 2nd nuclear localization signal (NLS#2) and the transmembrane domain. The EGFP coding region is nucleotides 3954-4682 and the *drm* NLS#1D5del mutant coding region is nucleotides 4683-5033. The resulting nucleic acid sequence is SEQ ID NO:8 and the amino acid sequence of the NLS#1D5del mutant is SEQ ID NO:33.

NLS#2D5del: EGFP-NLS#2 mutant nucleotide sequence (SEQ ID NO:6) was digested with BsrGI and Bpu1102I. The double stranded synthetic oligonucleotide:

GTACAAGTCCGGACTCAGAATGAGGGCTTCAGGCCTGAGTCT TACTCCCGAGT (SEQ ID NO:18)

was ligated into the digested plasmid producing an EGFP-DRM fusion minus the 1st and 2nd nuclear localization signals and the transmembrane domain. The EGFP coding region is nucleotides 3954-4682 and the DRM nls2\tm mutant coding region is nucleotides 4683-5033. The resulting nucleic acid is SEQ ID NO:9 and the amino acid sequence of the NLS#2D5del mutant is SEQ ID NO:34.

DAvaI: The EGFP-DRM nucleotide sequence (SEQ ID NO:1) was digested with AvaI and the synthetic ds oligonucleotide:

CCGGGGACGAGGACAGCTGTAATTA CCTGCTCCT GTC GACATTAATGGCC (SEQ ID NO:10)

was ligated in , introducing a stop codon at base 4878 in the EGFP/DRM sequence. The resulting nucleic acid sequence is SEQ ID NO:19 and the amino acid sequence of the DAvaI mutant is SEQ ID NO:35.

Although the present process has been described with reference to specific details of certain embodiments thereof, it is not intended that such details should be regarded as limitations upon the scope of the invention except as and to the extent that they are included in the accompanying claims.

Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

REFERENCES

1. **Athanasiou, M., G. Mavrothalassitis, C. C. Yuan, and D. G. Blair.** 1996. The *gag-myb-ets* fusion oncogene alters the apoptotic response and growth factor dependence of interleukin-3 dependent murine cells. Oncogene **12**:337-344.

2. **Barnes, J. L., and S. Milani.** 1995. In situ hybridization in the study of the kidney and renal diseases. Seminars in nephrology, v. 15, No. 1:9-28.

3. **Blair, D. G., M. A. Hull, and E. A. Finch.** 1979. The isolation and preliminary characterization of temperature sensitive transformation mutants of Moloney Sarcoma Virus. Virology **95:**303-316.

4. **Boowmeester, T., S. H. Kim, Y. Sasai, B. Lu, and E. M. De Robertis.** 1996. Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of spemann's organizer. Nature **382:**595-601.

5. **Boyd, J. M., S. Malstrom, T. Subramanian, L. K. Venkatesh, U. Schaeper, B. Elangovan, C. D'Sa-Eipper, and G. Chinnadurai.** 1994. Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set of cellular proteins. Cell **79:**341-351.

6. **Brody, J. S., and M. C. Williams.** 1992. Pulmonary alveolar epithelial cell differentiation. Ann. Rev. Physiol. **54:**351-371.

7. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156-159.

8. **Contente, S., K. Kenyon, D. Rimoldi, and R. M. Friedman.** 1990. Expression of gene rrg is associated with reversion of NIH3T3 transformed by LTR-c-H-ras. Science **249:**797-798.

9. **Denhardt, D. T.** 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. **23:**641-646.

10. **Enomoto, H., T. Ozaki, E. I. Takahashi, N. Nomura, S. Tabata, H. Takahashi, N. Ohnuma, M. Tanabe, J. Iwai, M. Yoshida, T. Matsunaga, and S. Sakiyama.** 1994. Identification of human DAN gene, mapping to the putative neuroblastoma tumor suppressor locus. Oncogene **9:**2785-2791.

11. Genetic Computer Group. 1994. Program manual for the Wisconsin GCG package. Version 8.0, University of Wisconsin, Madison.

12. **Gillet, G., M. Guerin, A. Trembleau, and G. Brun.** 1995. A BCL-2 related gene is activated in avian cells transformed by the Rous sarcoma virus. EMBO J. **14:**1372-1381.

13. **Glück, U., D. J. Kwiatkowski, and A. Ben-Ze'ev.** 1993. Suppression of tumorigenicity in simian virus 40-transformed 3T3 cells transfected with α-actinin cDNA. Proc. Natl. Acad. Sci. USA **90**:383-387.

14. **Gordon, J. I., and M. L. Hermiston.** 1994. Differentiation and self-renewal in the mouse gastrointestinal epithelium. Curr. Opin. Cell Biol. **6**:795-803.

15. **Gross-Bellard, M., P. Oudet, and P. Chambon.** 1973. Isolation of high-molecular-weight DNA from mammalian cells. Eur. J. Biochem. **36**:32-38.

16. **Gum, J. R., J. W. Hicks, N. W. Toribara, E-M. Rothe, R. E. Lagace, and Y. S. Kim.** 1992. The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. J. Biol. Chem. **267**:21375-21383.

17. **Hall, P. A., P. J. Coates, B. Ansam, and D. Hopwood.** 1994. Regulation of cell number in the mammalian gastrointestinal tract: the importance of apoptosis. J. Cell Sci. **107**:3569-3577.

18. **Hamelin, R., B. L. Brizzard, M. A. Nash, E. C. Murphy, and R. B. Arlinghaus.** 1985. Temperature-sensitive viral RNA expression in ts110 Moloney murine sarcoma virus-infected cells. J. Virol. **50**:478-488.

19. **Harada, H., M. Kitayawa, N. Tanaka, H. Yamamoto, K. Horada, M. Ishihara, and T. Taniguchi.** 1993. Anti-oncogenic and oncogenic potentials of interferon regulation factors-1 and -2. Science **259**:971-974.

20. **Houle, B., C. Rochette-Egly, and W.E.C. Bradley.** 1993. Tumor suppressive effect of the retinoic acid receptor β in human epidermoid lung cancer cells. Proc. Natl. Acad. Sci. USA **90**:985-989.

21. **Katzov, S., D. Martin-Zanca, and M. Barbacid.** 1989. Vav, a novel human oncogene derived from a locus ubiquitously expressed in hematopoietic cells. EMBO J. **8**:2283-2290.

22. **Kozak, M.** 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125-8133.

23. **Kozak, M.** 1992. Regulation of translation in eukaryotic systems. Ann. Rev. Cell Biol. **8**:197-225.

24. **Levine, A.** 1993. The tumor suppressor genes. Ann. Rev. Biochem. **62**:623-651.

25. **Liang, P., and A. B. Pardee.** 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science **247**:967-971.

26. **Lin, X., P. J. Nelson, B. Frankfort, E. Tombler, R. Johnson, and J. H. Gelman.** 1995. Isolation and characterization of a novel mitogenic regulatory gene, 322, which is transcriptionally suppressed in cells transformed by src and ras. Mol. Cell. Biol. **15**:2754-2762.

27. **Nuygen, M., P. E. Branton, P. A. Walton, Z. N. Oltvai, S. J. Korsmeyer, and G. C. Shore.** 1994. Role of membrane anchor domain of Bcl-2 in suppression of apoptosis caused by E1B-defective adenovirus. J. Biol. Chem. **269**:16521-16524.

28. **Ozaki, T., and S. Sakiyama.** 1993. Molecular cloning and characterization of a cDNA showing negative regulation in v-src-transformed 3Y1 rat fibroblasts. Proc. Natl. Acad. Sci. USA **90**:2593-2597.

29. **Ozaki, T., and S. Sakiyama.** 1994. Tumor-suppressive activity of NO3 gene product in v-src-transformed Rat 3Y1 fibroblasts. Cancer Res. **54**:646-648.

30. **Ozaki, T., Y. Nakamura, H. Enomoto, M. Hirose, and S. Sakiyama.** 1995. Overexpression of DAN gene product in normal rat fibroblasts causes a retardation of the entry into the S phase. Cancer Res. **55**:895-900.

31. **Prasad, G. L., R. A. Fuldner, and H. L. Cooper.** 1993. Expression of transduced tropomyosin 1 cDNA suppresses neoplastic growth of cells transformed by the ras oncogene. Proc. Natl. Acad. Sci. USA **90**:7039-7043.

32. **Preisig, P. A., and H. A. Franch.** 1995. Renal epithelial cell hyperplasia and hypertrophy. Seminars in nephrology **15(4)**:327-340.

33. **Rao, L., M. Debbas, P. Sabbatini, D. Hockenbery, S. Korsmeyer, and E. White.** 1992. The adenovirus E1A proteins induce apoptosis, which is inhibited by the E1B 19 kDa and Bcl-2 proteins. Proc. Natl. Acad. Sci. USA **89**:7742-7746.

34. **Sager, R.** 1989. Tumor suppressor genes: the puzzle and the promise. Science **246**:1406-1412.

35. **Sambrook, J., E. Fritsch, and T. Maniatis.** 1989. Molecular cloning: A laboratory manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

36. **Sanger, F.** 1981. Determination of nucleotide sequences in DNA. Science **214**:1205-1210.

37. **Sassoon, D., and N. Rosenthal.** 1993. Methods Enzymol. **225**:389-403.

38. **Shih, C., andR. A. Weinberg.** 1982. Isolation of a transforming sequence from a human bladder carcinoma cell line. Cell **29**:161-169.

39. **Sprague, J., J. H. Condra, H. Arnheiter, and R. A. Lazzarini.** 1983. Expression of a recombinant DNA gene coding for the resiculor stomatitis virus nucleocapsid protein. J. Virol. **45**: 773-781.

40. **Topol, L. Z., A. G. Tatosyan, D. Blair, and F. L. Kisselov.** 1991. A new recipient line for the transfection of biologically active oncogenes. Mol. Biol. (Translated) **25(2)**:541-551.

41. **Topol, L. Z., M. Marx, G. Calothy, and D. G. Blair.** 1995. Transformation-resistant mos revertant is unable to activate MAP kinase in response to v-mos or v-raf. Cell Growth Differ. **6**:27-38.

42. **Topol, L. Z., and D. G. Blair.** 1995. Activation of the mitogen-activated protein kinase cascade in response to the temperature inducible expression of v-mos kinase. Cell Growth Differ. **6**:1119-1127.

43. **White, E., P. Sabbatini, M. Debbas, W. S. M. Wold, D. I. Kusher, and L. Gooding.** 1992. The 19-kilodalton adenovirus E1B transforming protein inhibits programmed cell death and prevents cytolysis by tumor necrosis factor α. Mol. Cell. Biol. **12**:2570-2580.

44. **Zou, Z., A. Anisowicz, M.J.C. Hendrix, A. Thor, M. Neveu, S. Sheng, K. Rafidi, E. Seftor, and R. Sager.** 1994. Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. Science **263**:526-529.

45. **Lesser ML.** Design and implementation of clinical trials. In: Statistics in Medical Research - Methods and Issues with Applications in Cancer Research. Ed: Mike V and Stanley KF, New York, Wiley. 1982.

46. **Gehan EA, Schneiderman MA**: Experimental Design of Clinical Trials, in Holland JF and Frei E, Ill, eds. Cancer Medicine (2nd ed.). Lea and Febinger, Philadelphia, 531-553,1982.

47. **Gail M, Gart JJ**: The Determination of Sample Sizes for Use with the Exact Conditional Test in 2 x 2 Comparative Trials. Biometrics, 29, 441-448, 1973.

48. **Lee ET**: Statistical Methods for Survival Data Analysis, Wiley, New York, 1992.

49. **Kalbfleisch JD, Prentice RL**: The Statistical Analysis of Failure Time Data, New York, Wiley, 1980.

50. **Pastan et al.** "A retrovirus carrying an MDR1 cDNA confers multidrug resistance and polarized expression of P-glycoprotein in MDCK cells." *Proc. Nat. Acad. Sci.* 85:4486 (1988)

51. **Miller et al.** "Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production." *Mol. Cell Biol.* 6:2895 (1986).

52. **Mitani et al.** "Transduction of human bone marrow by adenoviral vector." *Human Gene Therapy* 5:941-948 (1994).

53. **Goodman et al.** "Recombinant adeno-associated virus-mediated gene transfer into hematopoietic progenitor cells." *Blood* 84:1492-1500 (1994)

54. **Naidini et al.** "*In vivo* gene delivery and stable transduction of nondividing cells by a lentiviral vector." *Science* 272:263-267 (1996))

55. **Agrawal et al.** "Cell-cycle kinetics and VSV-G pseudotyped retrovirus mediated gene transfer in blood-derived $CD34^+$ cells." *Exp. Hematol.* 24:738-747 (1996).

56. **Schwarzenberger et al.** "Targeted gene transfer to human hematopoietic progenitor cell lines through the *c-kit* receptor." *Blood* 87:472-478 (1996).

57. **Fields, *et al.*** (1990) Virology, Raven Press, New York.

58. **Michieli, P., Li, W., Lorenzi, M. V., Miki, T., Zakut, R., Givol, D., and Pierce, J. H.** (1996) *Oncogene* 12, 775-784.

59. **Crystal, R.G.** 1997. Phase I study of direct administration of a replication deficient adenovirus vector containing *E. coli* cytosine deaminase gene to metastatic colon carcinoma of the liver in association with the oral administration of the pro-drug 5-fluorocytosine. *Human Gene Therapy* 8:985-1001.

60. **Alvarez, R.D. and D.T. Curiel.** 1997. A phase I study of recombinant adenovirus vector-mediated delivery of an anti-erbB-2 single chain (sFv) antibody gene from previously treated ovarian and extraovarian cancer patients. *Hum. Gene Ther.* 8:229-242.

61. **Lewin,** "Genes V" Oxford University Press Chapter 7, pp. 171-174 (1994).

62. **Sambrook et al.**, *Molecular Cloning: A Laboratory Manual.* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989).

63. **Lewin**, "Genes V" Oxford University Press Chapter 1, pp. 9-13 (1994).

64. **Kunkel *et al.***, *Methods Enzymol.* 154:367 (1987).

65. **Topol, LZ, Marx, M, Laugier, D, Bogdanova, NN, Boubnov, NV, Clausen, PA, Calothy, G and Blair, DG.** 1997. Identification of drm, a novel gene whose expression is suppressed in transformed cells and which can inhibit growth of normal but not transformed cells in culture. *Mol. Cell Biol.* 17:4801-4810.

66. **Yonish-Rouach E, Resvitzky D, Lotem J, Sachs L, Kimchi A and Oren M.** 1991. Wild-type p53 induces apoptosis of myeloid leukemia cells that is inhibited by interleukin-6. Nature 352:345-347.

67. **Goldstein S.** 1990. Replicative senescence: the human fibroblast comes of age. Science 249:1129-1133.

68. **Schneider C, King RM and Philipson L.** 1988. Genes specifically expressed at growth arrest of mammalian cells. Cell 54:787-793.

69. **Del Sal G, Ruaro ME, Philipson L and Schneider C.** 1992. The growth arrest-specific gene gas1 is involved in growth suppression. Cell 70:595-607.

70. **Brancolini C, Bottega S and Schneider C.** 1992. Gas 2, a growth arrest-specific protein, is a component of the microfilament network system. Journal of Cell Biology 117:1251-1261.

71. **Sagesaka T, Boubnov N, Okuyama T, Paulus H and Sarkar N.** 1994. Deoxyribonucleic acid replication in fetal cells. American Journal of Obstetrics and Gynecology 170:468-473.

72. **Topol LZ, Marx M, Laugier D, Bogdanova NN, Boubnov NV, Clausen PA, Calothy G and Blair DG.** 1997. Identification of *drm*, a novel gene whose expression is suppressed in transformed cells and which can inhibit growth of normal but not transformed cells in culture. Molecular and Cellular Biology 17:4801-4810.

73. **Gaypay G, Schmitt K, Fizames C, Jones M, Vega-Ozarny N, Spillet D, Muselet D, Prud'Homme J-F, Dib C, Auffray C, Morisette J, Weissenbach J and Goodfellow PN.** 1996. A radiation hybrid map of the human genome. Human Molecular Genetics 5:339-346.

74. **Kozak M.** 1991. Structure features in eukaryotic mRNAs that modulate the initiation of translation. Journal of Biological Chemistry 266:19867-19870.

75. **Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morisette J and Weissenbach J.** 1996. A comprehensive genetic maps of the human genome based on 5,264 microsatellites. Nature 380:152-154.

76. **Pagano M, Tam SW, Theodoras AM, Beer-Romero P, Del Sal G, Chau V, Yew PR, Draetta GF and Rolfe M. 1995.** Role of the ubiquitin-proteosome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. Science 269:682-685.

77. **Alessandrini A, Chiaur DS and Pagano M.** 1992. Regulation of the cyclin-dependent kinase inhibitor p27 by degradation and phosphorylation. Leukemia 11:342-345.

78. **Koff A, Cross F, Fisher A, Schumacher J, Leguellee K, Philippe M and Roberts JM.** 1991. Human cyclin E, a new cyclin that interacts with two members of the CDC2 gene family. Cell 66:1217-1228.

79. **Forss-Petter S, Danielson P and Sutcliffe JG.** 1986. Neuron-specific Enolase: Complete structure of rat mRNA, multiple transcriptional start sites and evidence suggesting post-transcriptional control. Journal of Neuroscience Research 16:141-156

80. **Spector DL, Fu X-D and Maniatis T.** 1991. Associations between distinct pre-mRNA splicing components and the cell nucleus. EMBO J. 10:3467-3481.

81. **Huang S, Deerinch J, Ellisman M and Spector DL.** 1994. In vivo analysis of the stability and transport of nuclear Poly(A)$^+$ RNA. Journal of Cell Biology 126: 878-899.

82. **Stauber RH, Horie K, Carney P, Hudson EA, Tarasova NI, Gaitanaris GA and Pavlakis GN.** 1998. Development and applications of enhanced green fluorescent protein mutants. BioTechniques 24:462-471.

83. **Forss-Petter S, Danielson P and Sutcliffe JG.** 1986. Neuron-specific Enolase: Complete structure of rat mRNA, multiple transcriptional start sites and evidence suggesting post-transcriptional control. Journal of Neuroscience Research 16:141-156.

84. **Tohyama T, Lee VM-Y and Trojanovski J.** 1993. Co-expression of low molecular weight neurofilament protein and glial fibrillary acidic protein in established human glioma cell lines. American Journal of Pathology 142:883-892.

85. **Pagano M, Tam SW, Theodoras AM, Beer-Romero P, Del Sal G, Chau V, Yew PR, Draetta GF and Rolfe M.** 1995. Role of the ubiquitin-proteosome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. Science 269:682-685

86. **Rolfe M, Chin MI and Pagano M.** 1997. The ubiquitin-mediated proteolytic pathway as a therapeutic area. Journal of Molecular medicine 75:8-17.

87. **Lee M, Larner JM and Hamlin JL.** 1997. Cloning and characterization of Chinese hamster p53 cDNA. Gene 184:177-183.

88. **Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, Urdea MS, Valenzuela P, and Barr PJ** 1984. Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae,* PNAS 82:4642-4646.

89. **Sternsdorf, T., Jensen, K., Zuchner, D. and Will, H.** 1997. Cellular localization, expression, and structure of the nuclear dot protein 52. J. Cell Biol. 138: 435-448.

90. **Harlow and Lane**. Antibodies, A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988.

91. **Martin EW**: Remington's Pharmaceutical Sciences, latest ed., Mack Publishing Co., Easton, PA.

TABLE 1. *Drm* IS PREFERENTIALLY EXPRESSED IN TERMINALLY-DIFFERENTIATED CELLS

Tissue	Cell type	Proliferation/Differentiation
Brain	Neuron	none/terminally
	Glial	low/diff.
Kidney	Tubular epithelial	low/diff.
Lung	Type 1 epithelial	none/terminally
Intestine	Goblet	low/diff.
Spleen	Megakaryocyte	diff.

TABLE 2. DRM Expression in Normal and Malignant C ll Lin s

Normal Cell Lines	Screened Amount of Cell Lines	Amount With Positive Expression
Diploid fibroblasts	10	10
Normal astrocytes	1	1
Normal melanocytes	1	0
Normal mammary gland	1	0
HUVEC	1	0
Malignant Cell Lines		
Adenocarcinoma	21	0
Fibrosarcoma	3	0
Sarcoma	5	0
Melanoma	5	0
Carcinoma	10	0
Astrocytoma	1	0
Rhabdomyosarcoma	1	0

What is claimed is:

1. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:2.

2. An isolated polypeptide having the amino acid sequence of SEQ ID NO:36.

3. An isolated nucleic acid encoding the polypeptide of claim 2.

4. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:3.

5. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:4.

6. A fragment of DRM protein comprising the amino acid sequence encoded by nucleotides 4689 through 5243 of SEQ ID NO: 1.

7. An isolated nucleic acid encoding the amino acid sequence of claim 6.

8. A fragment of DRM protein comprising the amino acid sequence encoded by nucleotides 4683 through 5147 of SEQ ID NO: 5.

9. An isolated nucleic acid encoding the amino acid sequence of claim 8.

10. A fragment of DRM protein comprising the amino acid sequence encoded by nucleotides 1339 through 1815 of SEQ ID NO: 6.

11. An isolated nucleic acid encoding the amino acid sequence of claim 10.

12. A fragment of DRM protein comprising the amino acid sequence encoded by nucleotides 4683 through 5129 of SEQ ID NO: 7.

13. An isolated nucleic acid encoding the amino acid sequence of claim 12.

14. A fragment of DRM protein comprising the amino acid sequence encoded by nucleotides 4683 through 5033 of SEQ ID NO: 8.

15. An isolated nucleic acid encoding the amino acid sequence of claim 14.

16. A fragment of DRM protein comprising the amino acid sequence encoded by nucleotides 4689 through 5243 of SEQ ID NO: 19, wherein a stop codon is introduced at nucleotide 4878 of SEQ ID NO: 19.

17. An isolated nucleic acid encoding the amino acid sequence of claim 16.

18. A method of arresting the growth of a cell, comprising administering to the cell an effective amount of DRM protein or an active fragment thereof.

19. The method of claim 18, wherein the cell is *in vivo*.

20. The method of claim 18, wherein the cell is *ex vivo*.

21. A method of inhibiting tumor cell growth, comprising administering to a tumor cell an effective amount of DRM protein or an active fragment thereof.

22. The method of claim 21, wherein the tumor cell is *in vivo*.

23. A method of treating a hyperproliferative cell disorder in a subject diagnosed with a hyperproliferative disorder, comprising administering to the subject an effective amount of DRM protein, or an active fragment thereof, in a pharmaceutically acceptable carrier.

24. A method of arresting the growth of a cell, comprising administering to the cell an effective amount of a nucleic acid encoding a DRM protein or an active fragment thereof.

25. The method of claim 24, wherein the cell is *in vivo*.

26. The method of claim 24, wherein the cell is *ex vivo*.

27. A method of inhibiting tumor cell growth, comprising administering to a tumor cell an effective amount of a nucleic acid encoding a DRM protein, or an active fragment thereof.

28. The method of claim 27, wherein the tumor cell is *in vivo*.

29. A method of treating a hyperproliferative cell disorder in a subject diagnosed with a hyperproliferative disorder, comprising administering, to a cell of the subject, an effective amount of a nucleic acid encoding a DRM protein, or an active fragment thereof, under conditions whereby the nucleic acid can be expressed in the cell.

30. The method of claim 24, wherein the nucleic acid is in a virus.

31. The method of claim 24, wherein the nucleic acid is in a liposome.

32. A method of identifying a subject at risk of developing a hyperproliferative cell disorder, comprising measuring the amount of DRM protein in a cell of the subject, whereby an amount of DRM protein in a cell which is less than the amount of DRM protein in a cell of a normal subject identifies a subject at risk of developing a hyperproliferative cell disorder.

33. A method of identifying a subject at risk of developing a hyperproliferative cell disorder, comprising measuring the amount of nucleic acid encoding DRM in a cell of the subject, whereby an amount of nucleic acid encoding DRM in a cell which is less than the amount of nucleic acid encoding DRM in a cell of a normal subject identifies a subject at risk of developing a hyperproliferative cell disorder.

34. A fusion polypeptide comprising a DRM protein and a green fluorescent protein.

35. The fusion polypeptide of claim 34, wherein the DRM protein comprises the amino acid sequence encoded by nucleotides 4689 through 5243 of SEQ ID NO:1.

36. The fusion polypeptide of claim 34, wherein the DRM protein comprises the amino acid sequence encoded by nucleotides 4683 through 5147 of SEQ ID NO:5.

37. The fusion polypeptide of claim 34, wherein the DRM protein comprises the amino acid sequence encoded by nucleotides 1339 through 1815 of SEQ ID NO:6.

38. The fusion polypeptide of claim 34, wherein the DRM protein comprises the amino acid sequence encoded by nucleotides 4683 through 5129 of SEQ ID NO:7.

39. The fusion polypeptide of claim 34, wherein the DRM protein comprises the amino acid sequence encoded by nucleotides 4683 through 5033 of SEQ ID NO:8.

40. The fusion polypeptide of claim 34, wherein the DRM protein comprises the amino acid sequence encoded by nucleotides 4683 through 5033 of SEQ ID NO:9.

41. The fusion polypeptide of claim 34, wherein the DRM protein comprises the amino acid sequence encoded by nucleotides 4689 through 5243 of SEQ ID NO:19, wherein a stop codon is introduced at nucleotide 4878 of SEQ ID NO:19.

42. A green fluorescent protein having increased stability, comprising a fusion protein comprising a DRM protein amino acid sequence, or an active fragment thereof, linked to a green fluorescent protein amino acid sequence.

43. A method of producing a green fluorescent protein having increased stability, comprising the steps of amplifying DRM by PCR using forward primer CGGGATCCAGAATGAATCGCACGGCATAC (SEQ ID NO: 11) and reverse primer GCGGATCCTTAATCCAAGTCGATGGATATGC (SEQ ID NO:12), digesting the PCR product with BamH1 and EcoRI and ligating the digested product in frame into the pEGFP-C1 vector digested with BglII and EcoRI.

44. An isolated nucleic acid encoding the fusion polypeptide of claim 34.

45. An isolated nucleic acid encoding the fusion polypeptide of claim 35.

46. An isolated nucleic acid encoding the fusion polypeptide of claim 36.

47. An isolated nucleic acid encoding the fusion polypeptide of claim 37.

48. An isolated nucleic acid encoding the fusion polypeptide of claim 38.

49. An isolated nucleic acid encoding the fusion polypeptide of claim 39.

50. An isolated nucleic acid encoding the fusion polypeptide of claim 40.

51. An isolated nucleic acid encoding the fusion polypeptide of claim 41.

52. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:1

53. An isolated polypeptide having the amino acid of SEQ ID NO:29.

54. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:5.

55. An isolated polypeptide having the amino acid sequence of SEQ ID NO:30.

56. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:6.

57. An isolated polypeptide having the amino acid sequence of SEQ ID NO:31.

58. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:7.

59. An isolated polypeptide having the amino acid sequence of SEQ ID NO:32.

60. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:8.

61. An isolated polypeptide having the amino acid sequence of SEQ ID NO:33.

62. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:9.

63. An isolated polypeptide having the amino acid sequence of SEQ ID NO:34.

64. An isolated nucleic acid having the nucleotide sequence of SEQ ID NO:19.

65. An isolated polypeptide having the amino acid sequence of SEQ ID NO:35.

SEQUENCE LISTING

<110> Government of the United States of America

BLAIR, DONALD
CLAUSEN, PETER
TOPOL, LILIA
MARX, MARIA
CALOTHY, GEORGES

<120> METHODS AND COMPOSITIONS FOR DRM, A SECRETED PROTEIN WITH CELL GROWTH INHIBITING ACTIVITY

<130> 14014.0318/P

<150> 60/079,440
<151> 1998-03-26

<160> 38

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 5243
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 1

GATCCACCGG ATCTAGATAA CTGATCATAA TCAGCCATAC CACATTTGTA GAGGTTTTAC 60
TTGCTTTAAA AAACCTCCCA CACCTCCCCC TGAACCTGAA ACATAAAATG AATGCAATTG 120
TTGTTGTTAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA 180
ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 240
ATGTATCTTA ACGCGTAAAT TGTAAGCGTT AATATTTTGT TAAAATTCGC GTTAAATTTT 300
TGTTAAATCA GCTCATTTTT TAACCAATAG GCCGAAATCG GCAAAATCCC TTATAAATCA 360
AAAGAATAGA CCGAGATAGG GTTGAGTGTT GTTCCAGTTT GGAACAAGAG TCCACTATTA 420
AAGAACGTGG ACTCCAACGT CAAAGGGCGA AAAACCGTCT ATCAGGGCGA TGGCCCACTA 480
CGTGAACCAT CACCCTAATC AAGTTTTTTG GGGTCGAGGT GCCGTAAAGC ACTAAATCGG 540
AACCCTAAAG GGAGCCCCCG ATTTAGAGCT TGACGGGGAA AGCCGGCGAA CGTGGCGAGA 600
AAGGAAGGGA AGAAAGCGAA AGGAGCGGGC GCTAGGGCGC TGGCAAGTGT AGCGGTCACG 660
CTGCGCGTAA CCACCACACC CGCCGCGCTT AATGCGCCGC TACAGGGCGC GTCAGGTGGC 720
ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT 780
ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG 840
AGTCCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC 900
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT 960
GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT 1020
CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG CCCAGTTCCG 1080
CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA TGCAGAGGCC GAGGCCGCCT 1140
CGGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT TGGAGGCCTA GGCTTTTGCA 1200
AAGATCGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC 1260

GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA 1320
ATCGGCTGCT CTGATGCCGC CGTGTTCCGC TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG 1380
TCAAGACCGA CCTGTCCGGT GCCCTGAATG AACTGCAAGA CGAGGCAGCG CGGCTATCGT 1440
GGCTGGCCAC GACGGGCGTT CCTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCGGGAA 1500
GGGACTGGCT GCTATTGGGC GAAGTGCCGG GGCAGGATCT CCTGTCATCT CACCTTGCTC 1560
CTGCCGAGAA AGTATCCATC ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG 1620
CTACCTGCCC ATTCGACCAC CAAGCGAAAC ATCGCATCGA GCGAGCACGT ACTCGGATGG 1680
AAGCCGGTCT TGTCGATCAG GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG 1740
AACTGTTCGC CAGGCTCAAG GCGAGCATGC CCGACGGCGA GGATCTCGTC GTGACCCATG 1800
GCGATGCCTG CTTGCCGAAT ATCATGGTGG AAAATGGCCG CTTTTCTGGA TTCATCGACT 1860
GTGGCCGGCT GGGTGTGGCG GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG 1920
CTGAAGAGCT TGGCGGCGAA TGGGCTGACC GCTTCCTCGT GCTTTACGGT ATCGCCGCTC 1980
CCGATTCGCA GCGCATCGCC TTCTATCGCC TTCTTGACGA GTTCTTCTGA GCGGGACTCT 2040
GGGGTTCGAA ATGACCGACC AAGCGACGCC CAACCTGCCA TCACGAGATT TCGATTCCAC 2100
CGCCGCCTTC TATGAAAGGT TGGGCTTCGG AATCGTTTTC CGGGACGCCG GCTGGATGAT 2160
CCTCCAGCGC GGGGATCTCA TGCTGGAGTT CTTCGCCCAC CCTAGGGGGA GGCTAACTGA 2220
AACACGGAAG GAGACAATAC CGGAAGGAAC CCGCGCTATG ACGGCAATAA AAAGACAGAA 2280
TAAAACGCAC GGTGTTGGGT CGTTTGTTCA TAAACGCGGG GTTCGGTCCC AGGGCTGGCA 2340
CTCTGTCGAT ACCCCACCGA GACCCCATTG GGGCCAATAC GCCCGCGTTT CTTCCTTTTC 2400
CCCACCCCAC CCCCCAAGTT CGGGTGAAGG CCCAGGGCTC GCAGCCAACG TCGGGGCGGC 2460
AGGCCCTGCC ATAGCCTCAG GTTACTCATA TATACTTTAG ATTGATTTAA AACTTCATTT 2520
TTAATTTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCATGACCA AAATCCCTTA 2580
ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCGTAGAA AAGATCAAAG GATCTTCTTG 2640
AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA AAAAAACCAC CGCTACCAGC 2700
GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT CCGAAGGTAA CTGGCTTCAG 2760
CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTTCAA 2820
GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC CTGTTACCAG TGGCTGCTGC 2880
CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA CGATAGTTAC CGGATAAGGC 2940
GCAGCGGTCG GGCTGAACGG GGGGTTCGTG CACACAGCCC AGCTTGGAGC GAACGACCTA 3000
CACCGAACTG AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG 3060
AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA CGAGGGAGCT 3120
TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG TTTCGCCACC TCTGACTTGA 3180
GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC 3240
GGCCTTTTTA CGGTTCCTGG CCTTTTGCTG GCCTTTTGCT CACATGTTCT TTCCTGCGTT 3300
ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCATGCAT TAGTTATTAA TAGTAATCAA 3360
TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA CTTACGGTAA 3420
ATGGCCCGCC TGGCTGACCG CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG 3480
TTCCCATAGT AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT 3540
AAACTGCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC CCTATTGACG 3600
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA TGGGACTTTC 3660
CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC CATGGTGATG CGGTTTTGGC 3720
AGTACATCAA TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA 3780
TTGACGTCAA TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA 3840
ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 3900
GCAGAGCTGG TTTAGTGAAC CGTCAGATCC GCTAGCGCTA CCGGTCGCCA CCATGGTGAG 3960
CAAGGGCGAG GAGCTGTTCA CCGGGGTGGT GCCCATCCTG GTCGAGCTGG ACGGCGACGT 4020
AAACGGCCAC AAGTTCAGCG TGTCCGGCGA GGGCGAGGGC GATGCCACCT ACGGCAAGCT 4080
GACCCTGAAG TTCATCTGCA CCACCGGCAA GCTGCCCGTG CCCTGGCCCA CCCTCGTGAC 4140
CACCCTGACC TACGGCGTGC AGTGCTTCAG CCGCTACCCC GACCACATGA AGCAGCACGA 4200
CTTCTTCAAG TCCGCCATGC CCGAAGGCTA CGTCCAGGAG CGCACCATCT TCTTCAAGGA 4260
CGACGGCAAC TACAAGACCC GCGCCGAGGT GAAGTTCGAG GGCGACACCC TGGTGAACCG 4320
CATCGAGCTG AAGGGCATCG ACTTCAAGGA GGACGGCAAC ATCCTGGGGC ACAAGCTGGA 4380
GTACAACTAC AACAGCCACA ACGTCTATAT CATGGCCGAC AAGCAGAAGA ACGGCATCAA 4440
GGTGAACTTC AAGATCCGCC ACAACATCGA GGACGGCAGC GTGCAGCTCG CCGACCACTA 4500
CCAGCAGAAC ACCCCCATCG GCGACGGCCC CGTGCTGCTG CCCGACAACC ACTACCTGAG 4560
CACCCAGTCC GCCCTGAGCA AAGACCCCAA CGAGAAGCGC GATCACATGG TCCTGCTGGA 4620
GTTCGTGACC GCCGCCGGGA TCACTCTCGG CATGGACGAA CTGTACAAGT CCGGACTCAG 4680

ATCCAGAATG AATCGCACGG CATACACCGT AGGAGCTTTG CTTCTCCTCC TGGGAACCCT 4740
ACTGCCAGCA GCTGAAGGGA AAAAGAAAGG GTCCCAAGGA GCCATCCCAC CTCCTGACAA 4800
GGCTCAGCAC AATGACTCCG AGCAGACCCA GTCCCCACCA CAACCTGGCT CCAGGACCCG 4860
GGGGCGGGGC CAGGGGCGGG GCACCGCCAT GCCTGGAGAG GAGGTGCTTG AGTCCAGCCA 4920
AGAGGCCCTG CATGTGACAG AGCGCAAATA CCTGAAGCGA GATTGGTGCA AAACTCAGCC 4980
CCTGAAGCAG ACCATCCATG AGGAGGGCTG CAACAGCCGC ACTATCATCA ATCGCTTCTG 5040
TTACGGCCAG TGCAACTCCT TCTACATCCC CAGGCATATC CGAAAAGAGG AAGGCTCCTT 5100
TCAGTCTTGC TCCTTCTGCA AGCCCAAGAA ATTCACCACC ATGATGGTCA CACTCAACTG 5160
TCCTGAGCTA CAGCCACCCA CCAAGAAGAA AAGAGTCACA CGCGTGAAGC AGTGTCGTTG 5220
CATATCCATC GACTTGGATT AAG 5243

<210> 2
<211> 3319
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 2

GAAAGCGCAG GCCCCGAGGA CCCGCCGCAC TGACAGTATG AGCCGCACAG CCTACACGGT 60
GGGAGCCCTG CTTCTCCTCT TGGGGACCCT GCTGCCGGCT GCTGAAGGGA AAAAGAAAGG 120
GTCCCAAGGT GCCATCCCCC CGCCAGACAA GGCCCAGCAC AATGACTCAG AGCAGACTCA 180
GTCGCCCCAG CAGCCTGGCT CCAGGAACCG GGGGCGGGGC CAAGGGCGGG GCACTGCCAT 240
GCCCGGGGAG GAGGTGCTGG AGTCCAGCCA AGAGGCCCTG CATGTGACGG AGCGCAAATA 300
CCTGAAGCGA GACTGGTGCA AAACCCAGCC GCTTAAGCAG ACCATCCACG AGGAAGGCTG 360
CAACAGTCGC ACCATCATCA ACCGCTTCTG TTACGGCCAG TGCAACTCTT TCTACATCCC 420
CAGGCACATC CGGAAGGAGG AAGGTTCCTT TCAGTCCTGC TCCTTCTGCA AGCCCAAGAA 480
ATTCACTACC ATGATGGTCA CACTCAACTG CCCTGAACTA CAGCCACCTA CCAAGAAGAA 540
GAGAGTCACA CGTGTGAAGC AGTGTCGTTG CATATCCATC GATTTGGATT AAGCCAAATC 600
CAGGTGCACC CAGCATGTCC TAGGAATGCA GACCCAGGAA GTCCCAGACC TAAAACAACC 660
AGATTCTTAC TTGGCTTAAA CCTAGAGGCC AGAAGAACCC CCAGCTGCCT CCTGGCAGGA 720
GCCTGCTTGT GCGTAGTTCG TGTGCATGAG TGTGGATGGG TGCCTGTGGG TGTTTTTAGA 780
CACCAGAGAA AACACAGTCT CTGCTAGAGA GCACTTCCTA TTTTGTAAAC CTATCTGCTT 840
TAATGGGGAT GTACCAGAAA CCCACCTCAC CCCGGCTCAC ATCTAAAGGG GCGGGGCCGT 900
GGTCTGGTTC TGACTTTGTG TTTTTGTGCC CTCCTGGGGA CCAGAATCTC CTTTCGGAAT 960
GAATGTTCAT GGAAGAGGCT CCTCTGAGGG CAAGAGACCT GTTTTAGTGC TGCATTCGAC 1020
ATGGAAAAGT CCTTTTAACC TGTGCTTGCA TCCTCCTTTC CTCCTCCTCC TCACAATCCA 1080
TCTCTTCTTA AGTTGACAGT GACTATGTCA GTCTAATCTC TTGTTTGCCA GGGTTCCTAA 1140
ATTAATTCAC TTAACCATGA TGCAAATGTT TTTCATTTGG TGAAGACCTC CAGACTCTGG 1200
GAGAGGCTGG TGTGGGCAAG GACAAGCAGG ATAGTGGAGT GAGAAAGGGA GGGTGGAGGG 1260
TGAGGCCAAA TCAGGTCCAG CAAAAGTCAG TAGGGACATT GCAGAAGCTT GAAAGGCCAA 1320
TACCAGAACA CAGGCTGATG CTTCTGAGAA AGTCTTTTCC TAGTATTTAA CAAAACCCAA 1380
GTGAACAGAG GAGAAATGAG ATTGCCAGAA AGTGATTAAC TTTGGCCGTT GCAATCTGCT 1440
CAAACCTAAC ACCAAACTGA AAACATAAAT ACTGACCACT CCTATGTTCG GACCCAAGCA 1500
AGTTAGCTAA ACCAAACCAA CTCCTCTGCT TTGTCCCTCA GGTGGAAAAG AGAGGTAGTT 1560
TAGAACTCTC TGCATAGGGG TGGGAATTAA TCAAAAACCT CAGAGGCTGA AATTCCTAAT 1620
ACCTTTCCTT TATCGTGGTT ATAGTCAGCT CATTTCCATT CCACTATTTC CCATAATGCT 1680
TCTGAGAGCC ACTAACTTGA TTGATAAAGA TCCTGCCTCT GCTGAGTGTA CCTGACAGTA 1740
GTCTAAGATG AGAGAGTTTA GGGACTACTC TGTTTTAACA AGAAATATTT TGGGGGTCTT 1800
TTTGTTTTAA CTATTGTCAG GAGATTGGGC TAAAGAGAAG ACGACGAGAG TAAGGAAATA 1860
AAGGGAATTG CCTCTGGCTA GAGAGTAGTT AGGTGTTAAT ACCTGGTAGA GATGTAAGGG 1920
ATATGACCTC CCTTTCTTTA TGTGCTCACT TGAGGATCTG AGGGGACCCT GTTAGGAGAG 1980
CATAGCATCA TGATGTATTA GCTGTTCATC TGCTACTGGT TGGATGGACA TAACTATTGT 2040

AACTATTCAG TATTTACTGG TAGGCACTGT CCTCTGATTA AACTTGGCCT ACTGGCAATG 2100
GCTACTTAGG ATTGATCTAA GGGCCAAAGT GCAGGGTGGG TGAACTTTAT TGTACTTTGG 2160
ATTTGGTTAA CCTGTTTTCC TCAAGCCTGA GGTTTTATAT ACAAACTCCC TGAATACTCT 2220
TTTTGCCTTG TTACTTCTCA GCCTCCTAGC CAAGTCCTAT GTAATATGGA AAACAAACAC 2280
TGCAGACTTG AGATTCAGTT GCCGATCAAG GCTCTGGCAT TCAGAGAACC CTTGCAACTC 2340
GAGAAGCTGT TTTTGATTTC GTTTTTGTTT TGAACCGGTG CTCTCCCATC TAACAACTAA 2400
CSAGGACCAT TTCCAGGCGG GAGATATTTT AAACACCCAA AATGTTGGGT CTGATTTCCA 2460
AACTTTTAAA CTCACTACTG ATGATTCTCA CGCTAGGCGA ATTTGTCCAA ACACATAGTG 2520
TGTGTGTTTT GTATACACTG TATGACCCCA CCCCAAATCT TTGTATTGTC CACATTCTCC 2580
AACAATAAAG CACAGAGTGG ATTTAATTAA GCACACAAAT GCTAAGGCAG AATTTTGAGG 2640
GTGGGAGAGA AGAAAAGGGA AAGAAGCTGA AAATGTAAAA CCACACCAGG GAGGAAAAAT 2700
GACATTCAGA ACCACCAAAC ACTGAATTTC TCTTGTTGTT TTAACTCTSC CACAAGAATG 2760
CAWTTTCGTT AATGGAGATG ACTTAAGTTG GCAGCAGAAA TCTTCTTTTA GGAGCTTGTC 2820
CCCCAKTYTT GCACATAAGT GCAGATTTGC CCCAAGTAAA GAGAATTTCC TCAACACTAA 2880
CTTCACGGGG ATAATCACCA CCTAAMCRCC CTTAAAGCAW ATCACTAGCC AAAGAGGGGA 2940
ATATCTGTTC TTCTTACTGT GCCTATATTA AGACTAGTAC AAATGTGGTG TGTCTTCCAA 3000
CTTTCAKTGA AAATGCCATA TCTATACCAT ATTTTATTCG AGTCACTGAT GATGTAATGA 3060
TATATTTTTT CATTATTATA GTAGAATATT TTTATGGCAA GAWATTTGTG GTCTTGATCA 3120
TACCTATTAA AATAATGCCA AACACCAAAT ATGAATTTTA TGATGTACAC TTTGTGCTTG 3180
GCATTAAAAG ARAAAAACAC ACACCGGAAT TCCAGCTGAG CGCCGGTCGC TACCATTACC 3240
AGTTGGTCTG GTGTCAAAAG CCGAATTCTG CAGATATCCA TCACACTGGC GGCCGCTCGA 3300
GCATGCATCT AGAGGGCCC 3319

<210> 3
<211> 3795
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 3

GCGGCCGCGA GCTCTAATAC GACTCACTAT AGGGCGTCGA CTCGATCAGA TACATAGTAA 60
CCCAAGCTGA CACAAGCTTA GAACCTACAG TCGGAGCAGG AGTTGAATGT CACATTATCA 120
GCTCCAAACT TGAACCTGCT CCAAAGTATT AAGTTAATGT CAGAAAAACA ATGACATTTA 180
AGAATATTTT TAATGAAACA TTCAATTATC TTGGTTCGAT GCTAGCCTTA GGGTTGGATG 240
GCCCTCACTT GCCAGAAGTT GTCCTTTAAA GGAGATCCAT CTTAGGCTGC TTTTTGTCTC 300
TTAGAGATAA TTGGTCTAGA TAATGATACC AACTTGTCTG GTTCCTTGGA GATGAAGGTT 360
ATATTAAAAA GGTTATGTCA ATATGCACTT AGTGGTTGCC ACATGCAATA CTGGTATTCA 420
GCGGACAGAA AATGGATGCT TCCTTGCTGT TCTTGTGCAG CAAACCTTAA CCATGGGGCA 480
GAGGAAACCC CAGGGTAGCT GCCATGCCTG GAAGAGACAT TATGTATTTG AAACTGTTCT 540
CATTTGAAAA GAAAGCCTTC AATGCTTTAA TAACTCTTGG TGTGCCCCAG GCCAGCAAGT 600
GTTCCAGGCT TTTAGCTGGG TGGGAAGGCT GGCTGACTGA GTTAGGATCT TCATATTAAT 660
GCTTTCCCAG AGGACTGTGT CCAGGGATAC TGCCCCAGGA GAATCCTGAC AGCCTGCTGC 720
CTCTCTTTCC CTTTTCCGCC TGTCTGCCCT GTCTTTTCTG AACAACACCG CCTCTGAAAA 780
GTCTCCTCTT CTCTTATTTG CTTTGTTTAC CTCATGTTCC TGTCTCTGTA TGTTTCTTCT 840
CCCACCAGGT GGGAGATCAT GCTTAGACTT ATTGCTTTAT TTATTTATAA TGTATTTATT 900
TATAATTTAT TTATTTATTA AATGTTATAT GCCCTTGCCA TATACGAGTC ATATCAAGGT 960
CCACATTTGC TCACAGTTCA TTGGCATCAA TTCTATTCTT ATGAATTGAA ATATTCCCGT 1020
ACTTACTCTC TATTGTGCCC ATTTTTCTAC CTTACACACA CTCTCTCTTC TTCTTCTTTC 1080
TTCTTCTTCT TCTTCTTCTT CTTCTTCTTC TTCTTCTTCT TCTTCTTCTT CTTCTTCTTT 1140
TTCTTTCTTT CTTTCTTCTT CTTCTTCTTC TTCTTCTTCT TCTTCTTCTT CTTCTTCTTC 1200
TTCTTCTTTT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCC 1260
ACATGTGGCT TGAAAGCAGA AGGACTGTTT GGGGAAATGA CACAGTAAAG CAGCAGGGGG 1320

AGGCAAATGT GAACAAGGTG AGGTGACAGA TATGCATGAA AATCCACAAT GAAACTCCGT 1380
CTTGTACACC AACTTAAAAA TTAAAGCCAG AGAAATTAAA GACCTACCTG GTCAATTAAT 1440
CAGACAAAAA AAAATTCTAT TCATACATAC AGTCACATAG ATGGGTAATG TATTTTACCA 1500
CTTAGAAAGG TTGAAAAGTG GGGTCTGGAG AAATGGCTCA TCAGCTAAGA ACACTTTCTG 1560
TTCTTCCAAG CGTTCTGAGT TCAGTTGCCA GCACTCACAT TGGGGGCTCA CAACTGCCTA 1620
TAATTCCAGC TTTAGGAGTT CTGGGTGTTT TATTGCCCTC CCTAGGCACA CACACGGATT 1680
ACACAGACAC ACACACACAC ACACACACAC ACACACACAC ACACACAAGT TGTTATATCA 1740
TGGCAGAAAG AATGATACCA GCCATCTTTA TCCTCTTGGC CTTCCGTACA TCCCTCTTTT 1800
TAGGTTCTTT TTTTTTTTGA CAGGTTTCCT GGGCTTTTTC CAATACTGGA ACAGTGAAAA 1860
GTCTCATGTC AAATTCAAGG ATAAATACAG TTAAGTGAGC ATTAAAAAAA GTCACATGCA 1920
ATTGTGTCAG GAGCCAGTAA GGAATTCTAA TAGGAGCTGG TTCAAAAGAG AGACGGGTCC 1980
TGACTGAGTT TAAAGCTTGG CAAATTCACT GTGTGACCTG TGTCGAATTA CTCAGTTTGA 2040
TGGCTGAGAG AATAATGGAA ATAATAGTAT CTAATGGCTG GTGATACTGT TAGAAGTCAG 2100
TGCAACTGAA GTGTGTGTTG AGTACAGTGT GTTAAGTGTA ATTATTGATT TTTACTAAAT 2160
AACTTTCTTA TTGTCTGTGT CCCCCTCTCT TTGTCCTTTG TCTAGAATGA ATCGCACCGC 2220
ATACACTGTG GGAGCGTTGC TTCTCCTCCT GGGGACCCTA CTGCCAACAG CTGAGGGGAA 2280
AAAGAAAGGT TCCCAAGGAG CCATTCCGCC TCCTGACAAG GCTCAGCACA ATGACTCTGA 2340
GCAGACCCAG TCCCCACCAC AACCTGGCTC CAGGACCCGG GGGCGGGGCC AGGGGCGGGG 2400
CACCGCCATG CCTGGAGAGG AGGTGCTTGA GTCCAGCCAA GAGGCCCTGC ACGTGACAGA 2460
GCGCAAGTAT CTGAAGCGAG ATTGGTGCAA AACTCAGCCC CTGAAGCAGA CCATCCACGA 2520
GGAGGGCTGC AACAGCCGCA CTATCATCAA CCGCTTCTGT TATGGCCAGT GCAACTCCTT 2580
CTACATCCCC AGGCACATCC GAAAGGAGGA AGGGTCCTTT CAGTGTTGCT CCTTCTGCAA 2640
GCCCAAGAAG TTCACCACCA TGATGGTCAC ACTCAACTGT CCTGAGCTAC AGCCACCCAC 2700
CAAGAAGAAA AGGGTCACAC GCGTGAAGCA GTGCCGTTGC ATATCCATCG ACTTGGATTA 2760
AGTCAAAGCG GGCACATTCA GCCTGTCATA GCCATGCTGA GAGAGCCACA CCCAAACCAC 2820
CCGATTCCTA CTTGGCTTAA ACCTAGAGGC CAGAAGAACC AGCAGTTGCT TCCTGGCTGG 2880
AGGCTGCTTA TGCATAGTGT ATGCGCATGA GTGTGCATGG GTGCCTGTGG GTGTTTCCAA 2940
ACACCAGCCG GAAACAGCCT TTGCTAGAAG GCACTTCCTG TTACTCTGCT TCAGATGGTC 3000
GGAAATGCCC ACACCACTGG ACCCAAACAT CCACAGGGGC AGGGCTGTAG TTGGCTTTGT 3060
CATTGTGTTC CATGTGCCTC CTGGGCACCA GGATTTCACT TGAGAATGAA TACTAATGGG 3120
GGAGGTAACT CTGAGGGCTG CATTAGACTC GGAACTGTTC AGTGCTCGCC CTATGCTCCC 3180
ATAGCCCATC CCTTTCTTTG CTCTCCCTGA CATCTCAGTC GTAGCCCATG TTCCTAAATT 3240
AATTCACTTG ACCGCGGGTG TAAGTCTTTT GTCTTGTGAA GAACCTTCAG AATGTGGGGA 3300
GACACGTGGT GATGGCAAAC GGGACAGAGG ACTGACGCAG GAACGGTCAG GCTGAGGACC 3360
AGTCTGGGCC AGTGACATTC AGTAGTGAGA TGTCTAGAGT TTAAAAGTTG TTTCCCAAAA 3420
CAATATTAGT CTTGTTTTTA GCAAAAGGGT TTTCCTGATA TTTAAAAGAA CCCAGACACA 3480
CAGAGGAAAA ATATAATCAG CAAAAAAACA AAACAAAACA AAATAACACA AACAATAACA 3540
ACAACAACAA ACAAAAACCC AATTCTCTGT GCCAGCTTCT GTGACCTACT GATACTAGCT 3600
GTAACTGATA CTAGCTGTTA AGGGTGAAAT GCTGACCACT CCTGTTTTAA GAACCAAGTG 3660
AAATTAAAAA AGAAAATGTG GCCTCCTACT TTACTTTGCC TCTCTGAAGT ACAACTGAGA 3720
GCCTTGTTCA CTGGGGTAAG AGAAGGCAAA TCCTCCTAAG CTTAGTTTCG CTGGATTAAC 3780
ATTGCTTGTC CGCCG 3795

<210> 4
<211> 3820
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 4

ACCTGGGGAG CCAGAGCACC GCAGTAGCGC ACTTTCCTTC GTGTTCTTCC CGCGTCGAGC 60
CCGAGTGGCT CCGGCCGCGG TCGCACGCAA CGCCACGCGT CCACAGCGAA GGACTTGAGG 120
ATCCACTGAG GTGACAGAAT GAATCGCACG GCATACACCG TAGGAGCTTT GCTTCTCCTC 180

CTGGGAACCC TACTGCCAGC AGCTGAAGGG AAAAAGAAAG GGTCCCAAGG AGCCATCCCA 240
CCTCCTGACA AGGCTCAGCA CAATGACTCC GAGCAGACCC AGTCCCCACC ACAACCTGGC 300
TCCAGGACCC GGGGGCGGGG CCAGGGGCGG GGCACCGCCA TGCCTGGAGA GGAGGTGCTT 360
GAGTCCAGCC AAGAGGCCCT GCATGTGACA GAGCGCAAAT ACCTGAAGCG AGATTGGTGC 420
AAAACTCAGC CCCTGAAGCA GACCATCCAT GAGGAGGGCT GCAACAGCCG CACTATCATC 480
AATCGCTTCT GTTACGGCCA GTGCAACTCC TTCTACATCC CCAGGCATAT CCGAAAAGAG 540
GAAGGCTCCT TTCAGTCTTG CTCCTTCTGC AAGCCCAAGA AATTCACCAC CATGATGGTC 600
ACACTCAACT GTCCTGAGCT ACAGCCACCC ACCAAGAAGA AAAGAGTCAC ACGCGTGAAG 660
CAGTGTCGTT GCATATCCAT CGACTTGGAT TAAGTCAAAG GGGGCACATT CAGCCTGTCA 720
TAGCCATGCC GAGAGCCACA CCCAAACCAC CCGATTCCTA CTTGGCTTAA ACCTAGAGGC 780
CAGAAGAACC AGCAGTTGCT TCCTGGCTGG AGGCTGCTTA TGCATAGAGT ATGCGCATGA 840
GTGTGCATGG GTACCTGTGG GTGTTTCCAA ACACCAGCGG AAACAGCCTC TGCAGGAAGG 900
CACTTCCTGT TACTGTGCTT CAGATGGTCG GAAATGCTCA CACCACTGGA CCCAACACCA 960
CAGGGGCAGG GCTGTAGATG ACTTTGACCT TGTGTTCCAT TGGCCTCCTG GGCACCAGGA 1020
TTTCATTTGA GAATGAATAC TAACGGAGGA GGTAACTCTG AGGGCCGCAT TAGACTCGGA 1080
ACAGTTTGTT CGTGCTCTCC CACAACCCAT TCCTTTCTTT GCTCTCCCTG ACCTTAGTCC 1140
ATGTTCTTAA ATTAATTCAC TTGATGTGAG TGTAAATTTC TTTCGTCTTG TGAAGAACCT 1200
TCAGAGTGTG GGGAGACAAG TGATAAAGGC AAACAGAACA GGGGATTGAC ACAGGAGCAT 1260
TGAGACTGAG GACCAGTCTG GCCAGTGAAA TTCAGTAGCA AGATGTTCAG AGTTTAAAGA 1320
TTGTTCCCCC CCAAACAATA TGAGTCTTGT TTTAGCAAAG GGGCTTTACT GATATTTAAA 1380
AGAACCCAGA CAGACAGAGG AGAAATATAA TCAGCAAAAA AACCAATTCT CTGTGCCGGT 1440
ATCTGTGACC TACTGACAAT ATCTGTAATC CAATGTTAAG GGTGAAATAT TGACCACTTC 1500
TGTTTTAAGA ACCAAGTGAA AGGAAAAAAA AAATATGGCC TTCTACTTAC TTTGCCTCTC 1560
AGGAGGATGA CTGAGAGCGT TGTTCGCTAG GGTAAGAAAG ACAAAACCTC CTAGGCTTAG 1620
TTTTGCTGGA TTATCATTGC TTTCCCATCA TTCCTGAAAA AATGCTTCAG AGATGCAGAA 1680
CCTTCCAATA AAATCGTGCT TTTCTTGAGA CCATTTGCCA GTAAGGGTCA GTGTTAGACG 1740
AGAGAGCTGT CTGCTGCATG TGAGTTAGAC ATGTCTGGGG CTTCTTCTGT TTGGCTTTTG 1800
TTATAGGAGA GAACCAGAGA TGAGAGAGCT GATGAGAGAA CAGAGACAGA GAGAGAGAGG 1860
GCCAATCCCT TAGGGAAGCA CTAGGGTATA TTAACAGGCC ACCTACACCC AATGGATCTA 1920
TGTGACATTG TAATCATTAT GCCTACTATG GATGCTGTCC TCTGAATACA CATGGCTGCC 1980
CAATGTCTAC TTAGCATCTA TGTAAGGGCC CAGAGAAAGG TGACTGGGTC TTGGTACATT 2040
TTGGTTTGGC TAAGCAATAC TCTTTTAAGA CTGACATTCT AGCTATAAAT GCCCCAGATA 2100
CTTTTTTTGC CTTTTCCTCT CAGAGCGACT AGTCAAGTGA TATGTCATTT GGAAGGCAGA 2160
CATTCACTGC CCATCAAAGA TACCACAGTC AAAGAACCAT TGGGAGTAAA GAAACTTTTT 2220
GTTTTGGTCT AGCCCACCCG CCCATGTAAC ATCGAAACAG GAACCATATT ACAAGGCAAA 2280
AGCTATCTTG AATTCCCAAA ACACTGGGTC TAATTTTGAA AGTTTAAAAG TCACTGGTGA 2340
TGACTCCACA GTAAGTGAAC TTGTGCGAGC ATAGCCGTGA GTTTCATTTG TACTGCGTGC 2400
TCCTTCACTG AATCTTTGAG GCTTCCATAT CCATAGCCAC ATAGTCACAG GGTGGATTTG 2460
ATTAGGCCCA CACATACAAA GGTGGGTTTG GAGGGTGGTG AAGAGGGAAA AATAAGAGAG 2520
GATGAAGATG AAAATATAGA CCCACACCAG AGAGGAAAAA TGACCCTCGG TGCTGAAAAA 2580
CACTGTGTCC CATCTTAATT CTGCCACAAA CATGCAGTCT TGCTAAAAAT CAACAACAAC 2640
AATAATAAAA ATGTTTGGCA GCCACAGTTA CCTTTAGGAG CTTGTACCAC AGTCTCTCTT 2700
GTAAGCTGGA TTTAGATTTG GTTCTTGACG ATTGCCTCAA AATTAACTTC TTTGAAACGA 2760
TCAGCAGCAT AAGTGCCCTA AAAGCACATC ACTGGCCAAC GGCTGGGACG TCTGCCTTCC 2820
TTGCCGTGCC TAGATCAAGA CCATCAGAAA ATGTGTCCGC TGCCGTTTAT TGGAGATGCC 2880
CCGTCTGTCG CTGATTCTGG ACGCACCAGC GATGCAAGGA TGGACACTTT CTCCAACATT 2940
GTAGTAGAAC CAATTTTTTT TGGCAAGCTT TGTTGCAGTC TCCACCTTAC CTGTTAAATA 3000
ATGCCAGAAA CCAAATATGA ATCTTACGGC ATTCAATTGT GCTTGGCACT GAAAGAGGAA 3060
AGCCACACAC CAGATAAGTC TGAGTGCCCC TTTGCCATTG TACTCTTCAA AGTGAGAAAC 3120
CTGGAGGAAG GATAGTCTCC ATGTGGAATG TGAATAAGCA AAAGAGTTAT GGTTATTTAA 3180
TGTAATTAGG AATTCTAGGT CCTTCGGTTA CTGTGATTTC GAATGTTTTC TTTCTCTGTT 3240
TTATACGACA GCCTCTGAGT TGGGGCAAAG AAGAAACAGG CCGTTGTATG TTGCTAGAGA 3300
CTTTCGTCAG GTCAGGGGGA CACACAGTCT TGTCACATAT GAAGAGATGT TACCAAGTCA 3360
ACGACAAGCC TTATTTTTTA ACGTTGAATG TTCCTTAAAG GCTGACACTT CTGAAGCAAT 3420
GTTAGGAAAG ACTTTAAATG TTATTTTGAG AGACTTCTGT GCGTATACAA GCAGATAATG 3480
ACGGCATGTT CAGACAAGCA GAACATTTCT AAACGAGAAG TCCGAGCTGA ACGACTGAAA 3540
AGAGATTCCT CGCCATATTG AATATCATCT ACATTGTGTA TTTAATATAC TTTAATCATT 3600

TTGAAACAAC GAAGGATTAT GCAGGCTATG ACGGAACTAC TACCTTGCTA TGGATGAGGG 3660
TTGGGCAGGA TTTAATGGTC TCATAGAAGC TAATTTGGCT TAAAGTTTTA TGAATCTGTA 3720
ACTAGAATTT TATTTTCACC CTAATAACAT TCTATATAAC CTTTGCCAAA AAAGCAATCA 3780
ATAAATTAAC CTCTTCTTTC TGTGGCAAAA AAAAAAAAAA 3820

<210> 5
<211> 5168
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 5

GATCCACCGG ATCTAGATAA CTGATCATAA TCAGCCATAC CACATTTGTA GAGGTTTTAC 60
TTGCTTTAAA AAACCTCCCA CACCTCCCCC TGAACCTGAA ACATAAAATG AATGCAATTG 120
TTGTTGTTAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA 180
ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 240
ATGTATCTTA ACGCGTAAAT TGTAAGCGTT AATATTTTGT TAAAATTCGC GTTAAATTTT 300
TGTTAAATCA GCTCATTTTT TAACCAATAG GCCGAAATCG GCAAAATCCC TTATAAATCA 360
AAAGAATAGA CCGAGATAGG GTTGAGTGTT GTTCCAGTTT GGAACAAGAG TCCACTATTA 420
AAGAACGTGG ACTCCAACGT CAAAGGGCGA AAAACCGTCT ATCAGGGCGA TGGCCCACTA 480
CGTGAACCAT CACCCTAATC AAGTTTTTTG GGGTCGAGGT GCCGTAAAGC ACTAAATCGG 540
AACCCTAAAG GGAGCCCCCG ATTTAGAGCT TGACGGGGAA AGCCGGCGAA CGTGGCGAGA 600
AAGGAAGGGA AGAAAGCGAA AGGAGCGGGC GCTAGGGCGC TGGCAAGTGT AGCGGTCACG 660
CTGCGCGTAA CCACCACACC CGCCGCGCTT AATGCGCCGC TACAGGGCGC GTCAGGTGGC 720
ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT 780
ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG 840
AGTCCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC 900
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT 960
GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT 1020
CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG CCCAGTTCCG 1080
CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA TGCAGAGGCC GAGGCCGCCT 1140
CGGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT TGGAGGCCTA GGCTTTTGCA 1200
AAGATCGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC 1260
GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA 1320
ATCGGCTGCT CTGATGCCGC CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC GGTTCTTTTT 1380
GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAAG ACGAGGCAGC GCGGCTATCG 1440
TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC TGAAGCGGGA 1500
AGGGACTGGC TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC TCACCTTGCT 1560
CCTGCCGAGA AAGTATCCAT CATGGCTGAT GCAATGCGGC GGCTGCATAC GCTTGATCCG 1620
GCTACCTGCC CATTCGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG 1680
GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC 1740
GAACTGTTCG CCAGGCTCAA GGCGAGCATG CCCGACGGCG AGGATCTCGT CGTGACCCAT 1800
GGCGATGCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC GCTTTTCTGG ATTCATCGAC 1860
TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT 1920
GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT 1980
CCCGATTCGC AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC 2040
TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC ATCACGAGAT TTCGATTCCA 2100
CCGCCGCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT CCGGGACGCC GGCTGGATGA 2160
TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCA CCCTAGGGGG AGGCTAACTG 2220
AAACACGGAA GGAGACAATA CCGGAAGGAA CCCGCGCTAT GACGGCAATA AAAAGACAGA 2280
ATAAAACGCA CGGTGTTGGG TCGTTTGTTC ATAAACGCGG GGTTCGGTCC CAGGGCTGGC 2340
ACTCTGTCGA TACCCCACCG AGACCCCATT GGGGCCAATA CGCCCGCGTT TCTTCCTTTT 2400

```
CCCCACCCCA CCCCCCAAGT TCGGGTGAAG GCCCAGGGCT CGCAGCCAAC GTCGGGGCGG   2460
CAGGCCCTGC CATAGCCTCA GGTTACTCAT ATATACTTTA GATTGATTTA AAACTTCATT   2520
TTTAATTTAA AAGGATCTAG GTGAAGATCC TTTTTGATAA TCTCATGACC AAAATCCCTT   2580
AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT   2640
GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAAACCA CCGCTACCAG   2700
CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA   2760
GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC CACCACTTCA   2820
AGAACTCTGT AGCACCGCCT ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG   2880
CCAGTGGCGA TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG   2940
CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG CGAACGACCT   3000
ACACCGAACT GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA   3060
GAAAGGCGGA CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC   3120
TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG   3180
AGCGTCGATT TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG   3240
CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC TTTCCTGCGT   3300
TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCATGCA TTAGTTATTA ATAGTAATCA   3360
ATTACGGGGT CATTAGTTCA TAGCCCATAT ATGGAGTTCC GCGTTACATA ACTTACGGTA   3420
AATGGCCCGC CTGGCTGACC GCCCAACGAC CCCCGCCCAT TGACGTCAAT AATGACGTAT   3480
GTTCCCATAG TAACGCCAAT AGGGACTTTC CATTGACGTC AATGGGTGGA GTATTTACGG   3540
TAAACTGCCC ACTTGGCAGT ACATCAAGTG TATCATATGC CAAGTACGCC CCCTATTGAC   3600
GTCAATGACG GTAAATGGCC CGCCTGGCAT TATGCCCAGT ACATGACCTT ATGGGACTTT   3660
CCTACTTGGC AGTACATCTA CGTATTAGTC ATCGCTATTA CCATGGTGAT GCGGTTTTGG   3720
CAGTACATCA ATGGGCGTGG ATAGCGGTTT GACTCACGGG GATTTCCAAG TCTCCACCCC   3780
ATTGACGTCA ATGGGAGTTT GTTTTGGCAC CAAAATCAAC GGGACTTTCC AAAATGTCGT   3840
AACAACTCCG CCCCATTGAC GCAAATGGGC GGTAGGCGTG TACGGTGGGA GGTCTATATA   3900
AGCAGAGCTG GTTTAGTGAA CCGTCAGATC CGCTAGCGCT ACCGGTCGCC ACCATGGTGA   3960
GCAAGGGCGA GGAGCTGTTC ACCGGGGTGG TGCCCATCCT GGTCGAGCTG GACGGCGACG   4020
TAAACGGCCA CAAGTTCAGC GTGTCCGGCG AGGGCGAGGG CGATGCCACC TACGGCAAGC   4080
TGACCCTGAA GTTCATCTGC ACCACCGGCA AGCTGCCCGT GCCCTGGCCC ACCCTCGTGA   4140
CCACCCTGAC CTACGGCGTG CAGTGCTTCA GCCGCTACCC CGACCACATG AAGCAGCACG   4200
ACTTCTTCAA GTCCGCCATG CCCGAAGGCT ACGTCCAGGA GCGCACCATC TTCTTCAAGG   4260
ACGACGGCAA CTACAAGACC CGCGCCGAGG TGAAGTTCGA GGGCGACACC CTGGTGAACC   4320
GCATCGAGCT GAAGGGCATC GACTTCAAGG AGGACGGCAA CATCCTGGGG CACAAGCTGG   4380
AGTACAACTA CAACAGCCAC AACGTCTATA TCATGGCCGA CAAGCAGAAG AACGGCATCA   4440
AGGTGAACTT CAAGATCCGC CACAACATCG AGGACGGCAG CGTGCAGCTC GCCGACCACT   4500
ACCAGCAGAA CACCCCCATC GGCGACGGCC CCGTGCTGCT GCCCGACAAC CACTACCTGA   4560
GCACCCAGTC CGCCCTGAGC AAAGACCCCA ACGAGAAGCG CGATCACATG GTCCTGCTGG   4620
AGTTCGTGAC CGCCGCCGGG ATCACTCTCG GCATGGACGA ACTGTACAAG TCCGGACTCA   4680
GATCCAGAAT GAATCGCACG GCATACACCG TAGGAGCTTT GCTTCTCCTC CTGGGAACCC   4740
TACTGCCAGC AGCTGAAGGG AAAAAGAAAG GGTCCCAAGG AGCCATCCCA CCTCCTGACA   4800
AGGCTCAGCA CAATGACTCC GAGCAGACCC AGTCCCCACC ACAACCTGGC TCCAGGACCC   4860
GGGGGCGGGG CCAGGGGCGG GGCACCGCCA TGCCTGGAGA GGAGGTGCTT GAGTCCAGCC   4920
AAGAGGCCCT GCATGTGACA GAGCGCAAAT ACCTGAAGCG AGATTGGTGC AAAACTCAGC   4980
CCCTGAAGCA GACCATCCAT GAGGAGGGCT GCAACAGCCG CACTATCATC AATCGCTTCT   5040
GTTACGGCCA GTGCAACTCC TTCTACATCC CCAGGCATAT CCGAAAAGAG GAAGGCTCCT   5100
TTCAGTCTTG CTCCTTCTGC AAGCCCAAGA AATTCACCAC CATGTAAGTC GCTTCGACTT   5160
GGATTAAG                                                           5168
```

<210> 6
<211> 5166
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 6

TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 60
CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC CCCGCCCATT 120
GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA GGGACTTTCC ATTGACGTCA 180
ATGGGTGGAG TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC 240
AAGTACGCCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA 300
CATGACCTTA TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 360
CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG ACTCACGGGG 420
ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC AAAATCAACG 480
GGACTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT 540
ACGGTGGGAG GTCTATATAA GCAGAGCTGG TTTAGTGAAC CGTCAGATCC GCTAGCGCTA 600
CCGGTCGCCA CCATGGTGAG CAAGGGCGAG GAGCTGTTCA CCGGGGTGGT GCCCATCCTG 660
GTCGAGCTGG ACGGCGACGT AAACGGCCAC AAGTTCAGCG TGTCCGGCGA GGGCGAGGGC 720
GATGCCACCT ACGGCAAGCT GACCCTGAAG TTCATCTGCA CCACCGGCAA GCTGCCCGTG 780
CCCTGGCCCA CCCTCGTGAC CACCCTGACC TACGGCGTGC AGTGCTTCAG CCGCTACCCC 840
GACCACATGA AGCAGCACGA CTTCTTCAAG TCCGCCATGC CCGAAGGCTA CGTCCAGGAG 900
CGCACCATCT TCTTCAAGGA CGACGGCAAC TACAAGACCC GCGCCGAGGT GAAGTTCGAG 960
GGCGACACCC TGGTGAACCG CATCGAGCTG AAGGGCATCG ACTTCAAGGA GGACGGCAAC 1020
ATCCTGGGGC ACAAGCTGGA GTACAACTAC AACAGCCACA ACGTCTATAT CATGGCCGAC 1080
AAGCAGAAGA ACGGCATCAA GGTGAACTTC AAGATCCGCC ACAACATCGA GGACGGCAGC 1140
GTGCAGCTCG CCGACCACTA CCAGCAGAAC ACCCCCATCG GCGACGGCCC CGTGCTGCTG 1200
CCCGACAACC ACTACCTGAG CACCCAGTCC GCCCTGAGCA AAGACCCCAA CGAGAAGCGC 1260
GATCACATGG TCCTGCTGGA GTTCGTGACC GCCGCCGGGA TCACTCTCGG CATGGACGAG 1320
CTGTACAAGT CCGGACTCAG ATCTCGAGCT CAAGCTTCGA ATTCAATGAA TCGCACGGCA 1380
TACACCGTAG GAGCTTTGCT TCTCCTCCTG GGAACCCTAC TGCCAGCAGC TGAAGGGAAA 1440
AAGAAAGGGT CCCAAGGAGC CATCCCACCT CCTGACAAGG CTCAGCACAA TGACTCCGAG 1500
CAGACCCAGT CCCCACCACA ACCTGGCTCC AGGACCCGGG GGCGGGGCCA GGGGCGGGGC 1560
ACCGCCATGC CTGGAGAGGA GGTGCTTGAG TCCAGCCAAG AGGCCCTGCA TGTGACAGAG 1620
CGCAAATACC TGAAGCGAGA TTGGTGCAAA ACTCAGCCCC TGAAGCAGAC CATCCATGAG 1680
GAGGGCTGCA ACAGCCGCAC TATCATCAAT CGCTTCTGTT ACGGCCAGTG CAACTCCTTC 1740
TACATCCCCA GGCATATCCG AAAAGAGGAA GGCTCCTTTC AGTCTTGCTC CTTCTGCAAG 1800
CCCAAGATAT TCACCACCAT GTAAGGATCC ACCGGATCTA GATAACTGAT CATAATCAGC 1860
CATACCACAT TTGTAGAGGT TTTACTTGCT TTAAAAAACC TCCCACACCT CCCCCTGAAC 1920
CTGAAACATA AAATGAATGC AATTGTTGTT GTTAACTTGT TTATTGCAGC TTATAATGGT 1980
TACAAATAAA GCAATAGCAT CACAAATTTC ACAAATAAAG CATTTTTTTC ACTGCATTCT 2040
AGTTGTGGTT TGTCCAAACT CATCAATGTA TCTTAACGCG TAAATTGTAA GCGTTAATAT 2100
TTTGTTAAAA TTCGCGTTAA ATTTTTGTTA AATCAGCTCA TTTTTTAACC AATAGGCCGA 2160
AATCGGCAAA ATCCCTTATA AATCAAAAGA ATAGACCGAG ATAGGGTTGA GTGTTGTTCC 2220
AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC AACGTCAAAG GGCGAAAAAC 2280
CGTCTATCAG GGCGATGGCC CACTACGTGA ACCATCACCC TAATCAAGTT TTTTGGGGTC 2340
GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC CCCCGATTTA GAGCTTGACG 2400
GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA GCGAAAGGAG CGGGCGCTAG 2460
GGCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC ACACCCGCCG CGCTTAATGC 2520
GCCGCTACAG GGCGCGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG 2580
TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT 2640
GCTTCAATAA TATTGAAAAA GGAAGAGTCC TGAGGCGGAA AGAACCAGCT GTGGAATGTG 2700
TGTCAGTTAG GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT GCAAAGCATG 2760
CATCTCAATT AGTCAGCAAC CAGGTGTGGA AAGTCCCCAG GCTCCCCAGC AGGCAGAAGT 2820
ATGCAAAGCA TGCATCTCAA TTAGTCAGCA ACCATAGTCC CGCCCCTAAC TCCGCCCATC 2880
CCGCCCCTAA CTCCGCCCAG TTCCGCCCAT TCTCCGCCCC ATGGCTGACT AATTTTTTTT 2940
ATTTATGCAG AGGCCGAGGC CGCCTCGGCC TCTGAGCTAT TCCAGAAGTA GTGAGGAGGC 3000
TTTTTTGGAG GCCTAGGCTT TTGCAAAGAT CGATCAAGAG ACAGGATGAG GATCGTTTCG 3060
CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC GCTTGGGTGG AGAGGCTATT 3120
CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT GCCGCCGTGT TCCGGCTGTC 3180
AGCGCAGGGG CGCCCGGTTC TTTTTGTCAA GACCGACCTG TCCGGTGCCC TGAATGAACT 3240
GCAAGACGAG GCAGCGCGGC TATCGTGGCT GGCCACGACG GGCGTTCCTT GCGCAGCTGT 3300

GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA TTGGGCGAAG TGCCGGGGCA 3360
GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA TCCATCATGG CTGATGCAAT 3420
GCGGCGGCTG CATACGCTTG ATCCGGCTAC CTGCCCATTC GACCACCAAG CGAAACATCG 3480
CATCGAGCGA GCACGTACTC GGATGGAAGC CGGTCTTGTC GATCAGGATG ATCTGGACGA 3540
AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG CTCAAGGCGA GCATGCCCGA 3600
CGGCGAGGAT CTCGTCGTGA CCCATGGCGA TGCCTGCTTG CCGAATATCA TGGTGGAAAA 3660
TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC GCTATCAGGA 3720
CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC GGCGAATGGG CTGACCGCTT 3780
CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC ATCGCCTTCT ATCGCCTTCT 3840
TGACGAGTTC TTCTGAGCGG GACTCTGGGG TTCGAAATGA CCGACCAAGC GACGCCCAAC 3900
CTGCCATCAC GAGATTTCGA TTCCACCGCC GCCTTCTATG AAAGGTTGGG CTTCGGAATC 3960
GTTTTCCGGG ACGCCGGCTG GATGATCCTC CAGCGCGGGG ATCTCATGCT GGAGTTCTTC 4020
GCCCACCCTA GGGGGAGGCT AACTGAAACA CGGAAGGAGA CAATACCGGA AGGAACCCGC 4080
GCTATGACGG CAATAAAAAG ACAGAATAAA ACGCACGGTG TTGGGTCGTT TGTTCATAAA 4140
CGCGGGGTTC GGTCCCAGGG CTGGCACTCT GTCGATACCC CACCGAGACC CCATTGGGGC 4200
CAATACGCCC GCGTTTCTTC CTTTTCCCCA CCCCACCCCC CAAGTTCGGG TGAAGGCCCA 4260
GGGCTCGCAG CCAACGTCGG GGCGGCAGGC CCTGCCATAG CCTCAGGTTA CTCATATATA 4320
CTTTAGATTG ATTTAAAACT TCATTTTTAA TTTAAAAGGA TCTAGGTGAA GATCCTTTTT 4380
GATAATCTCA TGACCAAAAT CCCTTAACGT GAGTTTTCGT TCCACTGAGC GTCAGACCCC 4440
GTAGAAAAGA TCAAAGGATC TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG 4500
CAAACAAAAA AACCACCGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT 4560
CTTTTTCCGA AGGTAACTGG CTTCAGCAGA GCGCAGATAC CAAATACTGT CCTTCTAGTG 4620
TAGCCGTAGT TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG 4680
CTAATCCTGT TACCAGTGGC TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC 4740
TCAAGACGAT AGTTACCGGA TAAGGCGCAG CGGTCGGGCT GAACGGGGGG TTCGTGCACA 4800
CAGCCCAGCT TGGAGCGAAC GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA 4860
GAAAGCGCCA CGCTTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG CGGCAGGGTC 4920
GGAACAGGAG AGCGCACGAG GGAGCTTCCA GGGGGAAACG CCTGGTATCT TTATAGTCCT 4980
GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTTGT GATGCTCGTC AGGGGGGCGG 5040
AGCCTATGGA AAAACGCCAG CAACGCGGCC TTTTTACGGT TCCTGGCCTT TTGCTGGCCT 5100
TTTGCTCACA TGTTCTTTCC TGCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC 5160
ATGCAT 5166

<210> 7
<211> 5130
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 7

GATCCACCGG ATCTAGATAA CTGATCATAA TCAGCCATAC CACATTTGTA GAGGTTTTAC 60
TTGCTTTAAA AAACCTCCCA CACCTCCCCC TGAACCTGAA ACATAAAATG AATGCAATTG 120
TTGTTGTTAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA 180
ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 240
ATGTATCTTA ACGCGTAAAT TGTAAGCGTT AATATTTTGT TAAAATTCGC GTTAAATTTT 300
TGTTAAATCA GCTCATTTTT TAACCAATAG GCCGAAATCG GCAAAATCCC TTATAAATCA 360
AAAGAATAGA CCGAGATAGG GTTGAGTGTT GTTCCAGTTT GGAACAAGAG TCCACTATTA 420
AAGAACGTGG ACTCCAACGT CAAAGGGCGA AAAACCGTCT ATCAGGGCGA TGGCCCACTA 480
CGTGAACCAT CACCCTAATC AAGTTTTTTG GGGTCGAGGT GCCGTAAAGC ACTAAATCGG 540
AACCCTAAAG GGAGCCCCCG ATTTAGAGCT TGACGGGGAA AGCCGGCGAA CGTGGCGAGA 600
AAGGAAGGGA AGAAAGCGAA AGGAGCGGGC GCTAGGGCGC TGGCAAGTGT AGCGGTCACG 660
CTGCGCGTAA CCACCACACC CGCCGCGCTT AATGCGCCGC TACAGGGCGC GTCAGGTGGC 720

ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT 780
ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG 840
AGTCCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC 900
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT 960
GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT 1020
CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG CCCAGTTCCG 1080
CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA TGCAGAGGCC GAGGCCGCCT 1140
CGGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT TGGAGGCCTA GGCTTTTGCA 1200
AAGATCGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC 1260
GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA 1320
ATCGGCTGCT CTGATGCCGC CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC GGTTCTTTTT 1380
GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAAG ACGAGGCAGC GCGGCTATCG 1440
TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC TGAAGCGGGA 1500
AGGGACTGGC TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC TCACCTTGCT 1560
CCTGCCGAGA AAGTATCCAT CATGGCTGAT GCAATGCGGC GGCTGCATAC GCTTGATCCG 1620
GCTACCTGCC CATTCGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG 1680
GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC 1740
GAACTGTTCG CCAGGCTCAA GGCGAGCATG CCCGACGGCG AGGATCTCGT CGTGACCCAT 1800
GGCGATGCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC GCTTTTCTGG ATTCATCGAC 1860
TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT 1920
GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT 1980
CCCGATTCGC AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC 2040
TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC ATCACGAGAT TTCGATTCCA 2100
CCGCCGCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT CCGGGACGCC GGCTGGATGA 2160
TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCA CCCTAGGGGG AGGCTAACTG 2220
AAACACGGAA GGAGACAATA CCGGAAGGAA CCCGCGCTAT GACGGCAATA AAAAGACAGA 2280
ATAAAACGCA CGGTGTTGGG TCGTTTGTTC ATAAACGCGG GGTTCGGTCC CAGGGCTGGC 2340
ACTCTGTCGA TACCCCACCG AGACCCCATT GGGGCCAATA CGCCCGCGTT TCTTCCTTTT 2400
CCCCACCCCA CCCCCCAAGT TCGGGTGAAG GCCCAGGGCT CGCAGCCAAC GTCGGGGCGG 2460
CAGGCCCTGC CATAGCCTCA GGTTACTCAT ATATACTTTA GATTGATTTA AAACTTCATT 2520
TTTAATTTAA AAGGATCTAG GTGAAGATCC TTTTTGATAA TCTCATGACC AAAATCCCTT 2580
AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT 2640
GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAAACCA CCGCTACCAG 2700
CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA 2760
GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC CACCACTTCA 2820
AGAACTCTGT AGCACCGCCT ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG 2880
CCAGTGGCGA TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG 2940
CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG CGAACGACCT 3000
ACACCGAACT GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA 3060
GAAAGGCGGA CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC 3120
TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG 3180
AGCGTCGATT TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG 3240
CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC TTTCCTGCGT 3300
TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCATGCA TTAGTTATTA ATAGTAATCA 3360
ATTACGGGGT CATTAGTTCA TAGCCCATAT ATGGAGTTCC GCGTTACATA ACTTACGGTA 3420
AATGGCCCGC CTGGCTGACC GCCCAACGAC CCCCGCCCAT TGACGTCAAT AATGACGTAT 3480
GTTCCCATAG TAACGCCAAT AGGGACTTTC CATTGACGTC AATGGGTGGA GTATTTACGG 3540
TAAACTGCCC ACTTGGCAGT ACATCAAGTG TATCATATGC CAAGTACGCC CCCTATTGAC 3600
GTCAATGACG GTAAATGGCC CGCCTGGCAT TATGCCCAGT ACATGACCTT ATGGGACTTT 3660
CCTACTTGGC AGTACATCTA CGTATTAGTC ATCGCTATTA CCATGGTGAT GCGGTTTTGG 3720
CAGTACATCA ATGGGCGTGG ATAGCGGTTT GACTCACGGG GATTTCCAAG TCTCCACCCC 3780
ATTGACGTCA ATGGGAGTTT GTTTTGGCAC CAAAATCAAC GGGACTTTCC AAAATGTCGT 3840
AACAACTCCG CCCCATTGAC GCAAATGGGC GGTAGGCGTG TACGGTGGGA GGTCTATATA 3900
AGCAGAGCTG GTTTAGTGAA CCGTCAGATC CGCTAGCGCT ACCGGTCGCC ACCATGGTGA 3960
GCAAGGGCGA GGAGCTGTTC ACCGGGGTGG TGCCCATCCT GGTCGAGCTG GACGGCGACG 4020
TAAACGGCCA CAAGTTCAGC GTGTCCGGCG AGGGCGAGGG CGATGCCACC TACGGCAAGC 4080
TGACCCTGAA GTTCATCTGC ACCACCGGCA AGCTGCCCGT GCCCTGGCCC ACCCTCGTGA 4140

CCACCCTGAC CTACGGCGTG CAGTGCTTCA GCCGCTACCC CGACCACATG AAGCAGCACG 4200
ACTTCTTCAA GTCCGCCATG CCCGAAGGCT ACGTCCAGGA GCGCACCATC TTCTTCAAGG 4260
ACGACGGCAA CTACAAGACC CGCGCCGAGG TGAAGTTCGA GGGCGACACC CTGGTGAACC 4320
GCATCGAGCT GAAGGGCATC GACTTCAAGG AGGACGGCAA CATCCTGGGG CACAAGCTGG 4380
AGTACAACTA CAACAGCCAC AACGTCTATA TCATGGCCGA CAAGCAGAAG AACGGCATCA 4440
AGGTGAACTT CAAGATCCGC CACAACATCG AGGACGGCAG CGTGCAGCTC GCCGACCACT 4500
ACCAGCAGAA CACCCCCATC GGCGACGGCC CCGTGCTGCT GCCCGACAAC CACTACCTGA 4560
GCACCCAGTC CGCCCTGAGC AAAGACCCCA ACGAGAAGCG CGATCACATG GTCCTGCTGG 4620
AGTTCGTGAC CGCCGCCGGG ATCACTCTCG GCATGGACGA ACTGTACAAG TCCGGACTCA 4680
GAATGAGGGC TCAGCACAAT GACTCCGAGC AGACCCAGTC CCCACCACAA CCTGGCTCCA 4740
GGACCCGGGG GCGGGGCCAG GGGCGGGGCA CCGCCATGCC TGGAGAGGAG GTGCTTGAGT 4800
CCAGCCAAGA GGCCCTGCAT GTGACAGAGC GCAAATACCT GAAGCGAGAT TGGTGCAAAA 4860
CTCAGCCCCT GAAGCAGACC ATCCATGAGG AGGGCTGCAA CAGCCGCACT ATCATCAATC 4920
GCTTCTGTTA CGGCCAGTGC AACTCCTTCT ACATCCCCAG GCATATCCGA AAAGAGGAAG 4980
GCTCCTTTCA GTCTTGCTCC TTCTGCAAGC CCAAGAAATT CACCACCATG ATGGTCACAC 5040
TCAACTGTCC TGAGCTACAG CCACCCACCA AGAAGAAAAG AGTCACACGC GTGAAGCAGT 5100
GTCGTTGCAT ATCCATCGAC TTGGATTAAG 5130

<210> 8
<211> 5054
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 8

GATCCACCGG ATCTAGATAA CTGATCATAA TCAGCCATAC CACATTTGTA GAGGTTTTAC 60
TTGCTTTAAA AAACCTCCCA CACCTCCCCC TGAACCTGAA ACATAAAATG AATGCAATTG 120
TTGTTGTTAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA 180
ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 240
ATGTATCTTA ACGCGTAAAT TGTAAGCGTT AATATTTTGT TAAAATTCGC GTTAAATTTT 300
TGTTAAATCA GCTCATTTTT TAACCAATAG GCCGAAATCG GCAAAATCCC TTATAAATCA 360
AAAGAATAGA CCGAGATAGG GTTGAGTGTT GTTCCAGTTT GGAACAAGAG TCCACTATTA 420
AAGAACGTGG ACTCCAACGT CAAAGGGCGA AAAACCGTCT ATCAGGGCGA TGGCCCACTA 480
CGTGAACCAT CACCCTAATC AAGTTTTTTG GGGTCGAGGT GCCGTAAAGC ACTAAATCGG 540
AACCCTAAAG GGAGCCCCCG ATTTAGAGCT TGACGGGGAA AGCCGGCGAA CGTGGCGAGA 600
AAGGAAGGGA AGAAAGCGAA AGGAGCGGGC GCTAGGGCGC TGGCAAGTGT AGCGGTCACG 660
CTGCGCGTAA CCACCACACC CGCCGCGCTT AATGCGCCGC TACAGGGCGC GTCAGGTGGC 720
ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT 780
ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG 840
AGTCCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC 900
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT 960
GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT 1020
CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG CCCAGTTCCG 1080
CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA TGCAGAGGCC GAGGCCGCCT 1140
CGGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT TGGAGGCCTA GGCTTTTGCA 1200
AAGATCGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC 1260
GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA 1320
ATCGGCTGCT CTGATGCCGC CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC GGTTCTTTTT 1380
GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAAG ACGAGGCAGC GCGGCTATCG 1440
TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC TGAAGCGGGA 1500
AGGGACTGGC TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC TCACCTTGCT 1560
CCTGCCGAGA AAGTATCCAT CATGGCTGAT GCAATGCGGC GGCTGCATAC GCTTGATCCG 1620

GCTACCTGCC CATTCGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG 1680
GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC 1740
GAACTGTTCG CCAGGCTCAA GGCGAGCATG CCCGACGGCG AGGATCTCGT CGTGACCCAT 1800
GGCGATGCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC GCTTTTCTGG ATTCATCGAC 1860
TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT 1920
GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT 1980
CCCGATTCGC AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC 2040
TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC ATCACGAGAT TTCGATTCCA 2100
CCGCCGCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT CCGGGACGCC GGCTGGATGA 2160
TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCA CCCTAGGGGG AGGCTAACTG 2220
AAACACGGAA GGAGACAATA CCGGAAGGAA CCCGCGCTAT GACGGCAATA AAAAGACAGA 2280
ATAAAACGCA CGGTGTTGGG TCGTTTGTTC ATAAACGCGG GGTTCGGTCC CAGGGCTGGC 2340
ACTCTGTCGA TACCCCACCG AGACCCCATT GGGGCCAATA CGCCCGCGTT TCTTCCTTTT 2400
CCCCACCCCA CCCCCCAAGT TCGGGTGAAG GCCCAGGGCT CGCAGCCAAC GTCGGGGCGG 2460
CAGGCCCTGC CATAGCCTCA GGTTACTCAT ATATACTTTA GATTGATTTA AAACTTCATT 2520
TTTAATTTAA AAGGATCTAG GTGAAGATCC TTTTTGATAA TCTCATGACC AAAATCCCTT 2580
AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT 2640
GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAAACCA CCGCTACCAG 2700
CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA 2760
GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC CACCACTTCA 2820
AGAACTCTGT AGCACCGCCT ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG 2880
CCAGTGGCGA TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG 2940
CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG CGAACGACCT 3000
ACACCGAACT GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA 3060
GAAAGGCGGA CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC 3120
TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG 3180
AGCGTCGATT TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG 3240
CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC TTTCCTGCGT 3300
TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCATGCA TTAGTTATTA ATAGTAATCA 3360
ATTACGGGGT CATTAGTTCA TAGCCCATAT ATGGAGTTCC GCGTTACATA ACTTACGGTA 3420
AATGGCCCGC CTGGCTGACC GCCCAACGAC CCCCGCCCAT TGACGTCAAT AATGACGTAT 3480
GTTCCCATAG TAACGCCAAT AGGGACTTTC CATTGACGTC AATGGGTGGA GTATTTACGG 3540
TAAACTGCCC ACTTGGCAGT ACATCAAGTG TATCATATGC CAAGTACGCC CCCTATTGAC 3600
GTCAATGACG GTAAATGGCC CGCCTGGCAT TATGCCCAGT ACATGACCTT ATGGGACTTT 3660
CCTACTTGGC AGTACATCTA CGTATTAGTC ATCGCTATTA CCATGGTGAT GCGGTTTTGG 3720
CAGTACATCA ATGGGCGTGG ATAGCGGTTT GACTCACGGG GATTTCCAAG TCTCCACCCC 3780
ATTGACGTCA ATGGGAGTTT GTTTTGGCAC CAAAATCAAC GGGACTTTCC AAAATGTCGT 3840
AACAACTCCG CCCCATTGAC GCAAATGGGC GGTAGGCGTG TACGGTGGGA GGTCTATATA 3900
AGCAGAGCTG GTTTAGTGAA CCGTCAGATC CGCTAGCGCT ACCGGTCGCC ACCATGGTGA 3960
GCAAGGGCGA GGAGCTGTTC ACCGGGGTGG TGCCCATCCT GGTCGAGCTG GACGGCGACG 4020
TAAACGGCCA CAAGTTCAGC GTGTCCGGCG AGGGCGAGGG CGATGCCACC TACGGCAAGC 4080
TGACCCTGAA GTTCATCTGC ACCACCGGCA AGCTGCCCGT GCCCTGGCCC ACCCTCGTGA 4140
CCACCCTGAC CTACGGCGTG CAGTGCTTCA GCCGCTACCC CGACCACATG AAGCAGCACG 4200
ACTTCTTCAA GTCCGCCATG CCCGAAGGCT ACGTCCAGGA GCGCACCATC TTCTTCAAGG 4260
ACGACGGCAA CTACAAGACC CGCGCCGAGG TGAAGTTCGA GGGCGACACC CTGGTGAACC 4320
GCATCGAGCT GAAGGGCATC GACTTCAAGG AGGACGGCAA CATCCTGGGG CACAAGCTGG 4380
AGTACAACTA CAACAGCCAC AACGTCTATA TCATGGCCGA CAAGCAGAAG AACGGCATCA 4440
AGGTGAACTT CAAGATCCGC CACAACATCG AGGACGGCAG CGTGCAGCTC GCCGACCACT 4500
ACCAGCAGAA CACCCCCATC GGCGACGGCC CCGTGCTGCT GCCCGACAAC CACTACCTGA 4560
GCACCCAGTC CGCCCTGAGC AAAGACCCCA ACGAGAAGCG CGATCACATG GTCCTGCTGG 4620
AGTTCGTGAC CGCCGCCGGG ATCACTCTCG GCATGGACGA ACTGTACAAG TCCGGACTCA 4680
GAATGAGGGC TCAGCACAAT GACTCCGAGC AGACCCAGTC CCCACCACAA CCTGGCTCCA 4740
GGACCCGGGG GCGGGGCCAG GGGCGGGGCA CCGCCATGCC TGGAGAGGAG GTGCTTGAGT 4800
CCAGCCAAGA GGCCCTGCAT GTGACAGAGC GCAAATACCT GAAGCGAGAT TGGTGCAAAA 4860
CTCAGCCCCT GAAGCAGACC ATCCATGAGG AGGGCTGCAA CAGCCGCACT ATCATCAATC 4920
GCTTCTGTTA CGGCCAGTGC AACTCCTTCT ACATCCCCAG GCATATCCGA AAAGAGGAAG 4980
GCTCCTTTCA GTCTTGCTCC TTCTGCAAGC CCAAGAAATT CACCACCATG TAAGTCGCTT 5040

CGACTTGGAT TAAG 5054

<210> 9
<211> 5031
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 9

GATCCACCGG ATCTAGATAA CTGATCATAA TCAGCCATAC CACATTTGTA GAGGTTTTAC 60
TTGCTTTAAA AAACCTCCCA CACCTCCCCC TGAACCTGAA ACATAAAATG AATGCAATTG 120
TTGTTGTTAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA 180
ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 240
ATGTATCTTA ACGCGTAAAT TGTAAGCGTT AATATTTTGT TAAAATTCGC GTTAAATTTT 300
TGTTAAATCA GCTCATTTTT TAACCAATAG GCCGAAATCG GCAAAATCCC TTATAAATCA 360
AAAGAATAGA CCGAGATAGG GTTGAGTGTT GTTCCAGTTT GGAACAAGAG TCCACTATTA 420
AAGAACGTGG ACTCCAACGT CAAAGGGCGA AAAACCGTCT ATCAGGGCGA TGGCCCACTA 480
CGTGAACCAT CACCCTAATC AAGTTTTTTG GGGTCGAGGT GCCGTAAAGC ACTAAATCGG 540
AACCCTAAAG GGAGCCCCCG ATTTAGAGCT TGACGGGGAA AGCCGGCGAA CGTGGCGAGA 600
AAGGAAGGGA AGAAAGCGAA AGGAGCGGGC GCTAGGGCGC TGGCAAGTGT AGCGGTCACG 660
CTGCGCGTAA CCACCACACC CGCCGCGCTT AATGCGCCGC TACAGGGCGC GTCAGGTGGC 720
ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT 780
ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG 840
AGTCCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC 900
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT 960
GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT 1020
CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG CCCAGTTCCG 1080
CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA TGCAGAGGCC GAGGCCGCCT 1140
CGGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT TGGAGGCCTA GGCTTTTGCA 1200
AAGATCGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC 1260
GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA 1320
ATCGGCTGCT CTGATGCCGC CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC GGTTCTTTTT 1380
GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAAG ACGAGGCAGC GCGGCTATCG 1440
TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC TGAAGCGGGA 1500
AGGGACTGGC TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC TCACCTTGCT 1560
CCTGCCGAGA AAGTATCCAT CATGGCTGAT GCAATGCGGC GGCTGCATAC GCTTGATCCG 1620
GCTACCTGCC CATTCGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG 1680
GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC 1740
GAACTGTTCG CCAGGCTCAA GGCGAGCATG CCCGACGGCG AGGATCTCGT CGTGACCCAT 1800
GGCGATGCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC GCTTTTCTGG ATTCATCGAC 1860
TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT 1920
GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT 1980
CCCGATTCGC AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC 2040
TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC ATCACGAGAT TTCGATTCCA 2100
CCGCCGCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT CCGGGACGCC GGCTGGATGA 2160
TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCA CCCTAGGGGG AGGCTAACTG 2220
AAACACGGAA GGAGACAATA CCGGAAGGAA CCCGCGCTAT GACGGCAATA AAAAGACAGA 2280
ATAAAACGCA CGGTGTTGGG TCGTTTGTTC ATAAACGCGG GGTTCGGTCC CAGGGCTGGC 2340
ACTCTGTCGA TACCCCACCG AGACCCCATT GGGGCCAATA CGCCCGCGTT TCTTCCTTTT 2400
CCCCACCCCA CCCCCCAAGT TCGGGTGAAG GCCCAGGGCT CGCAGCCAAC GTCGGGGCGG 2460
CAGGCCCTGC CATAGCCTCA GGTTACTCAT ATATACTTTA GATTGATTTA AAACTTCATT 2520

```
TTTAATTTAA AAGGATCTAG GTGAAGATCC TTTTTGATAA TCTCATGACC AAAATCCCTT  2580
AACGTGAGTT TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT  2640
GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAAACCA CCGCTACCAG  2700
CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA  2760
GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC CACCACTTCA  2820
AGAACTCTGT AGCACCGCCT ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG  2880
CCAGTGGCGA TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG  2940
CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG CGAACGACCT  3000
ACACCGAACT GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA  3060
GAAAGGCGGA CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC  3120
TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG  3180
AGCGTCGATT TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG  3240
CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC TTTCCTGCGT  3300
TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCATGCA TTAGTTATTA ATAGTAATCA  3360
ATTACGGGGT CATTAGTTCA TAGCCCATAT ATGGAGTTCC GCGTTACATA ACTTACGGTA  3420
AATGGCCCGC CTGGCTGACC GCCCAACGAC CCCCGCCCAT TGACGTCAAT AATGACGTAT  3480
GTTCCCATAG TAACGCCAAT AGGGACTTTC CATTGACGTC AATGGGTGGA GTATTTACGG  3540
TAAACTGCCC ACTTGGCAGT ACATCAAGTG TATCATATGC CAAGTACGCC CCCTATTGAC  3600
GTCAATGACG GTAAATGGCC CGCCTGGCAT TATGCCCAGT ACATGACCTT ATGGGACTTT  3660
CCTACTTGGC AGTACATCTA CGTATTAGTC ATCGCTATTA CCATGGTGAT GCGGTTTTGG  3720
CAGTACATCA ATGGGCGTGG ATAGCGGTTT GACTCACGGG GATTTCCAAG TCTCCACCCC  3780
ATTGACGTCA ATGGGAGTTT GTTTTGGCAC CAAAATCAAC GGGACTTTCC AAAATGTCGT  3840
AACAACTCCG CCCCATTGAC GCAAATGGGC GGTAGGCGTG TACGGTGGGA GGTCTATATA  3900
AGCAGAGCTG GTTTAGTGAA CCGTCAGATC CGCTAGCGCT ACCGGTCGCC ACCATGGTGA  3960
GCAAGGGCGA GGAGCTGTTC ACCGGGGTGG TGCCCATCCT GGTCGAGCTG GACGGCGACG  4020
TAAACGGCCA CAAGTTCAGC GTGTCCGGCG AGGGCGAGGG CGATGCCACC TACGGCAAGC  4080
TGACCCTGAA GTTCATCTGC ACCACCGGCA AGCTGCCCGT GCCCTGGCCC ACCCTCGTGA  4140
CCACCCTGAC CTACGGCGTG CAGTGCTTCA GCCGCTACCC CGACCACATG AAGCAGCACG  4200
ACTTCTTCAA GTCCGCCATG CCCGAAGGCT ACGTCCAGGA GCGCACCATC TTCTTCAAGG  4260
ACGACGGCAA CTACAAGACC CGCGCCGAGG TGAAGTTCGA GGGCGACACC CTGGTGAACC  4320
GCATCGAGCT GAAGGGCATC GACTTCAAGG AGGACGGCAA CATCCTGGGG CACAAGCTGG  4380
AGTACAACTA CAACAGCCAC AACGTCTATA TCATGGCCGA CAAGCAGAAG AACGGCATCA  4440
AGGTGAACTT CAAGATCCGC CACAACATCG AGGACGGCAG CGTGCAGCTC GCCGACCACT  4500
ACCAGCAGAA CACCCCCATC GGCGACGGCC CCGTGCTGCT GCCCGACAAC CACTACCTGA  4560
GCACCCAGTC CGCCCTGAGC AAAGACCCCA ACGAGAAGCG CGATCACATG GTCCTGCTGG  4620
AGTTCGTGAC CGCCGCCGGG ATCACTCTCG GCATGGACGA ACTGTACAAG TCCGGACTCA  4680
GAATGAGGGC TCAGCACAAT GACTCCGAGC AGACCCAGTC CCCACCACAA CCTGGCTCCA  4740
GGACCCGGGG GCGGGGCCAG GGGCGGGGCA CCGCCATGCC TGGAGAGGAG GTGCTTGAGT  4800
CCAGCCAAGA GGCCCTGCAT GTGACAGAGC GCAAATACCT GAAGCGAGAT TGGTGCAAAA  4860
CTCAGCCCCT GAAGCAGACC ATCCATGAGG AGGGCTGCAA CAGCCGCACT ATCATCAATC  4920
GCTTCTGTTA CGGCCAGTGC AACTCCTTCT ACATCCCCAG GCATATCCGA AAAGAGGAAG  4980
GCTCCTTTCA GTCTTGCTCC TTCTGCAAGC CCAAGATATT CACCACCATG T            5031
```

<210> 10
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 10

```
CCGGGGACGA GGACAGCTGT AATTACCTGC TCCTGTCGAC ATTAATGGCC              50
```

<210> 11
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 11

CGGGATCCAG AATGAATCGC ACGGCATAC 29

<210> 12
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 12

GCGGATCCTT AATCCAAGTC GATGGATATG C 31

<210> 13
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 13

TAAGTCGCTT CGACGTACAT TCAGCGA 27

<210> 14
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 14

AGGAATTCAA TGAATCGCAC GGCATAC 27

<210> 15

<211> 32
<212> DNA
<213> Artificial S quence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 15

ACGGGATCCT TACATGGTGG TGAATACTTG GG 32

<210> 16
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 16

GTACAAGTCC GGACTCAGAA TGAGGGCTTC AGGCCTGAGT CTTACTCCCG AGT 53

<210> 17
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 17

GTACAAGTCC GGACTCAGAA TGAGGGCTTC AGGCCTGAGT CTTACTCCCG AGT 53

<210> 18
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 18

GTACAAGTCC GGACTCAGAA TGAGGGCTTC AGGCCTGAGT CTTACTCCCG AGT 53

<210> 19
<211> 5268

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 19

GATCCACCGG ATCTAGATAA CTGATCATAA TCAGCCATAC CACATTTGTA GAGGTTTTAC 60
TTGCTTTAAA AAACCTCCCA CACCTCCCCC TGAACCTGAA ACATAAAATG AATGCAATTG 120
TTGTTGTTAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA 180
ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 240
ATGTATCTTA ACGCGTAAAT TGTAAGCGTT AATATTTTGT TAAAATTCGC GTTAAATTTT 300
TGTTAAATCA GCTCATTTTT TAACCAATAG GCCGAAATCG GCAAAATCCC TTATAAATCA 360
AAAGAATAGA CCGAGATAGG GTTGAGTGTT GTTCCAGTTT GGAACAAGAG TCCACTATTA 420
AAGAACGTGG ACTCCAACGT CAAAGGGCGA AAAACCGTCT ATCAGGGCGA TGGCCCACTA 480
CGTGAACCAT CACCCTAATC AAGTTTTTTG GGGTCGAGGT GCCGTAAAGC ACTAAATCGG 540
AACCCTAAAG GGAGCCCCCG ATTTAGAGCT TGACGGGGAA AGCCGGCGAA CGTGGCGAGA 600
AAGGAAGGGA AGAAAGCGAA AGGAGCGGGC GCTAGGGCGC TGGCAAGTGT AGCGGTCACG 660
CTGCGCGTAA CCACCACACC CGCCGCGCTT AATGCGCCGC TACAGGGCGC GTCAGGTGGC 720
ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT 780
ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG 840
AGTCCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC 900
CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCAGGT 960
GTGGAAAGTC CCCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT 1020
CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG CCCAGTTCCG 1080
CCCATTCTCC GCCCCATGGC TGACTAATTT TTTTTATTTA TGCAGAGGCC GAGGCCGCCT 1140
CGGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT TGGAGGCCTA GGCTTTTGCA 1200
AAGATCGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC 1260
GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA 1320
ATCGGCTGCT CTGATGCCGC CGTGTTCCGC TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG 1380
TCAAGACCGA CCTGTCCGGT GCCCTGAATG AACTGCAAGA CGAGGCAGCG CGGCTATCGT 1440
GGCTGGCCAC GACGGGCGTT CCTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCGGGAA 1500
GGGACTGGCT GCTATTGGGC GAAGTGCCGG GGCAGGATCT CCTGTCATCT CACCTTGCTC 1560
CTGCCGAGAA AGTATCCATC ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG 1620
CTACCTGCCC ATTCGACCAC CAAGCGAAAC ATCGCATCGA GCGAGCACGT ACTCGGATGG 1680
AAGCCGGTCT TGTCGATCAG GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG 1740
AACTGTTCGC CAGGCTCAAG GCGAGCATGC CCGACGGCGA GGATCTCGTC GTGACCCATG 1800
GCGATGCCTG CTTGCCGAAT ATCATGGTGG AAAATGGCCG CTTTTCTGGA TTCATCGACT 1860
GTGGCCGGCT GGGTGTGGCG GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG 1920
CTGAAGAGCT TGGCGGCGAA TGGGCTGACC GCTTCCTCGT GCTTTACGGT ATCGCCGCTC 1980
CCGATTCGCA GCGCATCGCC TTCTATCGCC TTCTTGACGA GTTCTTCTGA GCGGGACTCT 2040
GGGGTTCGAA ATGACCGACC AAGCGACGCC CAACCTGCCA TCACGAGATT TCGATTCCAC 2100
CGCCGCCTTC TATGAAAGGT TGGGCTTCGG AATCGTTTTC CGGGACGCCG GCTGGATGAT 2160
CCTCCAGCGC GGGGATCTCA TGCTGGAGTT CTTCGCCCAC CCTAGGGGGA GGCTAACTGA 2220
AACACGGAAG GAGACAATAC CGGAAGGAAC CCGCGCTATG ACGGCAATAA AAAGACAGAA 2280
TAAAACGCAC GGTGTTGGGT CGTTTGTTCA TAAACGCGGG GTTCGGTCCC AGGGCTGGCA 2340
CTCTGTCGAT ACCCCACCGA GACCCCATTG GGGCCAATAC GCCCGCGTTT CTTCCTTTTC 2400
CCCACCCCAC CCCCCAAGTT CGGGTGAAGG CCCAGGGCTC GCAGCCAACG TCGGGGCGGC 2460
AGGCCCTGCC ATAGCCTCAG GTTACTCATA TATACTTTAG ATTGATTTAA AACTTCATTT 2520
TTAATTTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCATGACCA AAATCCCTTA 2580
ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCGTAGAA AAGATCAAAG GATCTTCTTG 2640
AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA AAAAAACCAC CGCTACCAGC 2700
GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT CCGAAGGTAA CTGGCTTCAG 2760
CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTTCAA 2820

GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC CTGTTACCAG TGGCTGCTGC 2880
CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA CGATAGTTAC CGGATAAGGC 2940
GCAGCGGTCG GGCTGAACGG GGGGTTCGTG CACACAGCCC AGCTTGGAGC GAACGACCTA 3000
CACCGAACTG AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG 3060
AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA CGAGGGAGCT 3120
TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG TTTCGCCACC TCTGACTTGA 3180
GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC 3240
GGCCTTTTTA CGGTTCCTGG CCTTTTGCTG GCCTTTTGCT CACATGTTCT TTCCTGCGTT 3300
ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCATGCAT TAGTTATTAA TAGTAATCAA 3360
TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA CTTACGGTAA 3420
ATGGCCCGCC TGGCTGACCG CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG 3480
TTCCCATAGT AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT 3540
AAACTGCCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC CCTATTGACG 3600
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA TGGGACTTTC 3660
CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC CATGGTGATG CGGTTTTGGC 3720
AGTACATCAA TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA 3780
TTGACGTCAA TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA 3840
ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 3900
GCAGAGCTGG TTTAGTGAAC CGTCAGATCC GCTAGCGCTA CCGGTCGCCA CCATGGTGAG 3960
CAAGGGCGAG GAGCTGTTCA CCGGGGTGGT GCCCATCCTG GTCGAGCTGG ACGGCGACGT 4020
AAACGGCCAC AAGTTCAGCG TGTCCGGCGA GGGCGAGGGC GATGCCACCT ACGGCAAGCT 4080
GACCCTGAAG TTCATCTGCA CCACCGGCAA GCTGCCCGTG CCCTGGCCCA CCCTCGTGAC 4140
CACCCTGACC TACGGCGTGC AGTGCTTCAG CCGCTACCCC GACCACATGA AGCAGCACGA 4200
CTTCTTCAAG TCCGCCATGC CCGAAGGCTA CGTCCAGGAG CGCACCATCT TCTTCAAGGA 4260
CGACGGCAAC TACAAGACCC GCGCCGAGGT GAAGTTCGAG GGCGACACCC TGGTGAACCG 4320
CATCGAGCTG AAGGGCATCG ACTTCAAGGA GGACGGCAAC ATCCTGGGGC ACAAGCTGGA 4380
GTACAACTAC AACAGCCACA ACGTCTATAT CATGGCCGAC AAGCAGAAGA ACGGCATCAA 4440
GGTGAACTTC AAGATCCGCC ACAACATCGA GGACGGCAGC GTGCAGCTCG CCGACCACTA 4500
CCAGCAGAAC ACCCCCATCG GCGACGGCCC CGTGCTGCTG CCCGACAACC ACTACCTGAG 4560
CACCCAGTCC GCCCTGAGCA AAGACCCCAA CGAGAAGCGC GATCACATGG TCCTGCTGGA 4620
GTTCGTGACC GCCGCCGGGA TCACTCTCGG CATGGACGAA CTGTACAAGT CCGGACTCAG 4680
ATCCAGAATG AATCGCACGG CATACACCGT AGGAGCTTTG CTTCTCCTCC TGGGAACCCT 4740
ACTGCCAGCA GCTGAAGGGA AAAAGAAAGG GTCCCAAGGA GCCATCCCAC CTCCTGACAA 4800
GGCTCAGCAC AATGACTCCG AGCAGACCCA GTCCCCACCA CAACCTGGCT CCAGGACCCG 4860
GGGACGAGGA CAGCTGTAAT TACCGGGGGC GGGGCCAGGG GCGGGGCACC GCCATGCCTG 4920
GAGAGGAGGT GCTTGAGTCC AGCCAAGAGG CCCTGCATGT GACAGAGCGC AAATACCTGA 4980
AGCGAGATTG GTGCAAAACT CAGCCCCTGA AGCAGACCAT CCATGAGGAG GGCTGCAACA 5040
GCCGCACTAT CATCAATCGC TTCTGTTACG GCCAGTGCAA CTCCTTCTAC ATCCCCAGGC 5100
ATATCCGAAA AGAGGAAGGC TCCTTTCAGT CTTGCTCCTT CTGCAAGCCC AAGAAATTCA 5160
CCACCATGAT GGTCACACTC AACTGTCCTG AGCTACAGCC ACCCACCAAG AAGAAAAGAG 5220
TCACACGCGT GAAGCAGTGT CGTTGCATAT CCATCGACTT GGATTAAG 5268

<210> 20
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 20

TCATTACATC ATCAGTGACT CG 22

<210> 21

<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 21

CAGATTTGGC TCAAGTAAAG AG 22

<210> 22
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 22

AGCCAGCGAA 10

<210> 23
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 23

GACCGCTTGT 10

<210> 24
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 24

AGGTGACCGT 10

<210> 25
<211> 10
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 25

GGTACTCCAC 10

<210> 26
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 26

GTTGCGATCC 10

<210> 27
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 27

CCGCTCGAGG TGACAGAATG AATCGC 26

<210> 28
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note =
synthetic construct

<400> 28

CCCGTTAACT TAGGCGTAGT CGGGCACGTC GTAGGGGTAA TCCAAGTCGA T 51

<210> 29
<211> 429
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

Ser gly leu arg ser

<400> 29

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
1 5 10 15
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
35 40 45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50 55 60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65 70 75 80
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85 90 95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100 105 110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115 120 125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130 135 140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145 150 155 160
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
165 170 175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180 185 190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
195 200 205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210 215 220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225 230 235 240
Gly Leu Arg Ser Arg Met Asn Arg Thr Ala Tyr Thr Val Gly Ala Leu
245 250 255
Leu Leu Leu Leu Gly Thr Leu Leu Pro Ala Ala Glu Gly Lys Lys Lys
260 265 270
Gly Ser Gln Gly Ala Ile Pro Pro Pro Asp Lys Ala Gln His Asn Asp
275 280 285
Ser Glu Gln Thr Gln Ser Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly
290 295 300
Arg Gly Gln Gly Arg Gly Thr Ala Met Pro Gly Glu Glu Val Leu Glu
305 310 315 320
Ser Ser Gln Glu Ala Leu His Val Thr Glu Arg Lys Tyr Leu Lys Arg
325 330 335
Asp Trp Cys Lys Thr Gln Pro Leu Lys Gln Thr Ile His Glu Glu Gly
340 345 350
Cys Asn Ser Arg Thr Ile Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn
355 360 365
Ser Phe Tyr Ile Pro Arg His Ile Arg Lys Glu Glu Gly Ser Phe Gln
370 375 380
Ser Cys Ser Phe Cys Lys Pro Lys Lys Phe Thr Thr Met Met Val Thr
385 390 395 400

```
Leu Asn Cys Pro Glu Leu Gln Pro Pro Thr Lys Lys Lys Arg Val Thr
                405                 410                 415
Arg Val Lys Gln Cys Arg Cys Ile Ser Ile Asp Leu Asp
            420                 425
```

<210> 30
<211> 397
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 30

```
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1               5                  10                  15
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
            20                  25                  30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
        35                  40                  45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
    50                  55                  60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65                  70                  75                  80
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
                85                  90                  95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
            100                 105                 110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
        115                 120                 125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
    130                 135                 140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145                 150                 155                 160
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
                165                 170                 175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
            180                 185                 190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
        195                 200                 205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
    210                 215                 220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225                 230                 235                 240
Gly Leu Arg Ser Arg Met Asn Arg Thr Ala Tyr Thr Val Gly Ala Leu
                245                 250                 255
Leu Leu Leu Leu Gly Thr Leu Leu Pro Ala Ala Glu Gly Lys Lys Lys
            260                 265                 270
Gly Ser Gln Gly Ala Ile Pro Pro Pro Asp Lys Ala Gln His Asn Asp
        275                 280                 285
Ser Glu Gln Thr Gln Ser Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly
    290                 295                 300
Arg Gly Gln Gly Arg Gly Thr Ala Met Pro Gly Glu Glu Val Leu Glu
305                 310                 315                 320
```

```
Ser Ser Gln Glu Ala Leu His Val Thr Glu Arg Lys Tyr Leu Lys Arg
                325                 330                 335
Asp Trp Cys Lys Thr Gln Pro Leu Lys Gln Thr Ile His Glu Glu Gly
            340                 345                 350

Cys Asn Ser Arg Thr Ile Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn
        355                 360                 365

Ser Phe Tyr Ile Pro Arg His Ile Arg Lys Glu Glu Gly Ser Phe Gln
    370                 375                 380
Ser Cys Ser Phe Cys Lys Pro Lys Lys Phe Thr Thr Met
385                 390                 395
```

<210> 31
<211> 403
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 31

```
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1               5                  10                  15
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
            20                  25                  30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
        35                  40                  45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
    50                  55                  60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65                  70                  75                  80
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
                85                  90                  95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
            100                 105                 110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
        115                 120                 125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
    130                 135                 140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145                 150                 155                 160
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
                165                 170                 175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
            180                 185                 190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
        195                 200                 205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
    210                 215                 220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225                 230                 235                 240
Gly Leu Arg Ser Arg Ala Gln Ala Ser Asn Ser Met Asn Arg Thr Ala
```

```
                245                 250                 255
Tyr Thr Val Gly Ala Leu Leu Leu Leu Leu Gly Thr Leu Leu Pro Ala
            260                 265                 270
Ala Glu Gly Lys Lys Lys Gly Ser Gln Gly Ala Ile Pro Pro Pro Asp
        275                 280                 285
Lys Ala Gln His Asn Asp Ser Glu Gln Thr Gln Ser Pro Pro Gln Pro
    290                 295                 300
Gly Ser Arg Thr Arg Gly Arg Gly Gln Gly Arg Gly Thr Ala Met Pro
305                 310                 315                 320
Gly Glu Glu Val Leu Glu Ser Ser Gln Glu Ala Leu His Val Thr Glu
                325                 330                 335

Arg Lys Tyr Leu Lys Arg Asp Trp Cys Lys Thr Gln Pro Leu Lys Gln
            340                 345                 350
Thr Ile His Glu Glu Gly Cys Asn Ser Arg Thr Ile Ile Asn Arg Phe
        355                 360                 365
Cys Tyr Gly Gln Cys Asn Ser Phe Tyr Ile Pro Arg His Ile Arg Lys
    370                 375                 380
Glu Glu Gly Ser Phe Gln Ser Cys Ser Phe Cys Lys Pro Lys Ile Phe
385                 390                 395                 400
Thr Thr Met
```

<210> 32
<211> 391
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 32

```
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1               5                  10                  15
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
            20                  25                  30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
        35                  40                  45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
    50                  55                  60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65                  70                  75                  80
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
                85                  90                  95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
            100                 105                 110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
        115                 120                 125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
    130                 135                 140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145                 150                 155                 160
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
                165                 170                 175
```

```
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
            180                 185                 190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
        195                 200                 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
    210                 215                 220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225                 230                 235                 240
Gly Leu Arg Met Arg Ala Gln His Asn Asp Ser Glu Gln Thr Gln Ser
                245                 250                 255
Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly Arg Gly Gln Gly Arg Gly
            260                 265                 270
Thr Ala Met Pro Gly Glu Glu Val Leu Glu Ser Ser Gln Glu Ala Leu
        275                 280                 285
His Val Thr Glu Arg Lys Tyr Leu Lys Arg Asp Trp Cys Lys Thr Gln
    290                 295                 300
Pro Leu Lys Gln Thr Ile His Glu Glu Gly Cys Asn Ser Arg Thr Ile
305                 310                 315                 320
Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn Ser Phe Tyr Ile Pro Arg
                325                 330                 335
His Ile Arg Lys Glu Glu Gly Ser Phe Gln Ser Cys Ser Phe Cys Lys
            340                 345                 350
Pro Lys Lys Phe Thr Thr Met Met Val Thr Leu Asn Cys Pro Glu Leu
        355                 360                 365
Gln Pro Pro Thr Lys Lys Lys Arg Val Thr Arg Val Lys Gln Cys Arg
    370                 375                 380
Cys Ile Ser Ile Asp Leu Asp
385                 390
```

<210> 33
<211> 359
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 33

```
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1               5                  10                  15
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
            20                  25                  30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
        35                  40                  45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
    50                  55                  60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65                  70                  75                  80
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
                85                  90                  95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
            100                 105                 110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
```

```
        115                 120                 125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
    130                 135                 140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145                 150                 155                 160
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
                165                 170                 175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
            180                 185                 190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
        195                 200                 205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
    210                 215                 220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225                 230                 235                 240
Gly Leu Arg Met Arg Ala Gln His Asn Asp Ser Glu Gln Thr Gln Ser
                245                 250                 255
Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly Arg Gly Gln Gly Arg Gly
            260                 265                 270
Thr Ala Met Pro Gly Glu Glu Val Leu Glu Ser Ser Gln Glu Ala Leu
        275                 280                 285
His Val Thr Glu Arg Lys Tyr Leu Lys Arg Asp Trp Cys Lys Thr Gln
    290                 295                 300
Pro Leu Lys Gln Thr Ile His Glu Glu Gly Cys Asn Ser Arg Thr Ile
305                 310                 315                 320
Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn Ser Phe Tyr Ile Pro Arg
                325                 330                 335
His Ile Arg Lys Glu Glu Gly Ser Phe Gln Ser Cys Ser Phe Cys Lys
            340                 345                 350
Pro Lys Lys Phe Thr Thr Met
        355
```

<210> 34
<211> 359
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 34

```
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1               5                  10                  15
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
            20                  25                  30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
        35                  40                  45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
    50                  55                  60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65                  70                  75                  80
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
                85                  90                  95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
```

```
            100                 105                 110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
        115                 120                 125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
    130                 135                 140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145                 150                 155                 160
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
                165                 170                 175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
            180                 185                 190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
        195                 200                 205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
    210                 215                 220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225                 230                 235                 240
Gly Leu Arg Met Arg Ala Gln His Asn Asp Ser Glu Gln Thr Gln Ser
                245                 250                 255
Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly Arg Gly Gln Gly Arg Gly
            260                 265                 270
Thr Ala Met Pro Gly Glu Glu Val Leu Glu Ser Ser Gln Glu Ala Leu
        275                 280                 285
His Val Thr Glu Arg Lys Tyr Leu Lys Arg Asp Trp Cys Lys Thr Gln
    290                 295                 300
Pro Leu Lys Gln Thr Ile His Glu Glu Gly Cys Asn Ser Arg Thr Ile
305                 310                 315                 320
Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn Ser Phe Tyr Ile Pro Arg
                325                 330                 335
His Ile Arg Lys Glu Glu Gly Ser Phe Gln Ser Cys Ser Phe Cys Lys
            340                 345                 350
Pro Lys Ile Phe Thr Thr Met
        355
```

<210> 35
<211> 308
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 35

```
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1                   5                  10                  15
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
            20                  25                  30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
        35                  40                  45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
    50                  55                  60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65                  70                  75                  80
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
```

```
                85                  90                  95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
            100                 105                 110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
        115                 120                 125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
    130                 135                 140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145                 150                 155                 160
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
                165                 170                 175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
            180                 185                 190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
        195                 200                 205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
    210                 215                 220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225                 230                 235                 240
Gly Leu Arg Ser Arg Met Asn Arg Thr Ala Tyr Thr Val Gly Ala Leu
                245                 250                 255
Leu Leu Leu Leu Gly Thr Leu Leu Pro Ala Ala Glu Gly Lys Lys Lys
            260                 265                 270
Gly Ser Gln Gly Ala Ile Pro Pro Pro Asp Lys Ala Gln His Asn Asp
        275                 280                 285
Ser Glu Gln Thr Gln Ser Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly
    290                 295                 300
Arg Gly Gln Leu
305
```

<210> 36
<211> 184
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 36

```
Met Ser Arg Thr Ala Tyr Thr Val Gly Ala Leu Leu Leu Leu Leu Gly
 1               5                  10                  15
Thr Leu Leu Pro Ala Ala Glu Gly Lys Lys Lys Gly Ser Gln Gly Ala
            20                  25                  30
Ile Pro Pro Pro Asp Lys Ala Gln His Asn Asp Ser Glu Gln Thr Gln
        35                  40                  45
Ser Pro Gln Gln Pro Gly Ser Arg Asn Arg Gly Arg Gly Gln Gly Arg
    50                  55                  60
Gly Thr Ala Met Pro Gly Glu Glu Val Leu Glu Ser Ser Gln Glu Ala
65                  70                  75                  80
Leu His Val Thr Glu Arg Lys Tyr Leu Lys Arg Asp Trp Cys Lys Thr
                85                  90                  95
Gln Pro Leu Lys Gln Thr Ile His Glu Glu Gly Cys Asn Ser Arg Thr
            100                 105                 110
Ile Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn Ser Phe Tyr Ile Pro
```

```
        115                 120                 125
Arg His Ile Arg Lys Glu Glu Gly Ser Phe Gln Ser Cys Ser Phe Cys
    130                 135                 140
Lys Pro Lys Lys Phe Thr Thr Met Met Val Thr Leu Asn Cys Pro Glu
145                 150                 155                 160
Leu Gln Pro Pro Thr Lys Lys Lys Arg Val Thr Arg Val Lys Gln Cys
                165                 170                 175
Arg Cys Ile Ser Ile Asp Leu Asp
            180
```

<210> 37
<211> 184
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 37

```
Met Asn Arg Thr Ala Tyr Thr Val Gly Ala Leu Leu Leu Leu Leu Gly
 1               5                  10                  15
Thr Leu Leu Pro Thr Ala Glu Gly Lys Lys Lys Gly Ser Gln Gly Ala
            20                  25                  30
Ile Pro Pro Pro Asp Lys Ala Gln His Asn Asp Ser Glu Gln Thr Gln
        35                  40                  45
Ser Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly Arg Gly Gln Gly Arg
    50                  55                  60
Gly Thr Ala Met Pro Gly Glu Glu Val Leu Glu Ser Ser Gln Glu Ala
65                  70                  75                  80
Leu His Val Thr Glu Arg Lys Tyr Leu Lys Arg Asp Trp Cys Lys Thr
                85                  90                  95
Gln Pro Leu Lys Gln Thr Ile His Glu Glu Gly Cys Asn Ser Arg Thr
            100                 105                 110
Ile Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn Ser Phe Tyr Ile Pro
        115                 120                 125
Arg His Ile Arg Lys Glu Glu Gly Ser Phe Gln Ser Cys Ser Phe Cys
    130                 135                 140
Lys Pro Lys Lys Phe Thr Thr Met Met Val Thr Leu Asn Cys Pro Glu
145                 150                 155                 160
Leu Gln Pro Pro Thr Lys Lys Lys Arg Val Thr Arg Val Lys Gln Cys
                165                 170                 175
Arg Cys Ile Ser Ile Asp Leu Asp
            180
```

<210> 38
<211> 184
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/Note = synthetic construct

<400> 38

```
Met Asn Arg Thr Ala Tyr Thr Val Gly Ala Leu Leu Leu Leu Leu Gly
 1               5                  10                  15
Thr Leu Leu Pro Ala Ala Glu Gly Lys Lys Lys Gly Ser Gln Gly Ala
            20                  25                  30
Ile Pro Pro Pro Asp Lys Ala Gln His Asn Asp Ser Glu Gln Thr Gln
        35                  40                  45
Ser Pro Pro Gln Pro Gly Ser Arg Thr Arg Gly Arg Gly Gln Gly Arg
    50                  55                  60
Gly Thr Ala Met Pro Gly Glu Glu Val Leu Glu Ser Ser Gln Glu Ala
65                  70                  75                  80
Leu His Val Thr Glu Arg Lys Tyr Leu Lys Arg Asp Trp Cys Lys Thr
                85                  90                  95
Gln Pro Leu Lys Gln Thr Ile His Glu Glu Gly Cys Asn Ser Arg Thr
            100                 105                 110
Ile Ile Asn Arg Phe Cys Tyr Gly Gln Cys Asn Ser Phe Tyr Ile Pro
        115                 120                 125
Arg His Ile Arg Lys Glu Glu Gly Ser Phe Gln Ser Cys Ser Phe Cys
    130                 135                 140
Lys Pro Lys Lys Phe Thr Thr Met Met Val Thr Leu Asn Cys Pro Glu
145                 150                 155                 160
Leu Gln Pro Pro Thr Lys Lys Lys Arg Val Thr Arg Val Lys Gln Cys
                165                 170                 175
Arg Cys Ile Ser Ile Asp Leu Asp
            180
```

INTERNATIONAL SEARCH REPORT

Intern ıal Application No
PCT/US 99/06675

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/475 C07K14/435 C12N15/62 C07K19/00
C12Q1/68 C07K16/18 G01N33/50 G01N33/53 A61K38/18
A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category '	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TOPOL L.Z. ET AL.: "Identification of drm, a novel gene whose expression is suppressed in transformed cells and which can inhibit growth of normal but not transformed cells in culture" MOLECULAR AND CELLULAR BIOLOGY, vol. 17, no. 8, 1 August 1997 (1997-08-01), pages 4801-4810, XP002066577 ISSN: 0270-7306 cited in the application	1-15,18, 20,24, 26,30,31
Y	the whole document	34-40, 42-50, 52-63

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

' Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 August 1999	09/09/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Macchia, G

1

Form PCT/ISA/210 (second sheet) (July 1992)

06/18/2003, EAST Version: 1.03.0002

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 97 26333 A (UNIV FLORIDA RES FOUND INC; ZOLOTUKHIN S.; MUZYCZKA N.; HAUSWIRTH W.W.) 24 July 1997 (1997-07-24) page 13, line 19-23 page 49, line 14 - page 50, line 22 page 86, line 8-12	34-40, 42-50, 52-63
P,X	WO 98 37195 A (REGENERON PHARM INC(US); UNIV CALIFORNIA; VALENZUELA; ECONOMIDES ET AL) 27 August 1998 (1998-08-27) abstract page 3, line 1-17 page 4, line 18-30 page 6, line 12-31 page 10, line 25-27 page 11, line 20-27 page 12, line 3-8 page 14, line 23 - page 15, line 30 Seq.ID:1,2 page 19 - page 20 page 25, line 18 - page 26, line 22; example 5 page 33 - page 35; claims	1-15, 18-31,34
P,X	WO 98 33918 A (THE REGENTS OF UNIVERSITY OF CALIFORNIA; HARLAND RICHARD; HSU DAVID) 6 August 1998 (1998-08-06) abstract page 2, line 24 - page 3, line 24 page 5, line 10-26 page 9, line 24-30 Seq.ID:1 page 12 - page 13 page 18; claims	1-15,34
P,X	HSU D.R. ET AL.: "The Xenopus dorsalizing factor Gremlin identifies a novel family of secretes proteins that antagonize BMP activities" MOLECULAR CELL, vol. 1, no. 5, April 1998 (1998-04), pages 673-683, XP002113640 page 676; figure 3A page 681, left-hand column, paragraph 5 page 683, right-hand column, paragraph 2	1-15
A	GARCIA-BUSTOS J. ET AL.: "Nuclear protein localization" BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1071, 1991, pages 83-101, XP002113641 page 86; table I	34-40, 42-50, 52-63

-/--

1

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/US 99/06675

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BOULIKAS T.: "Nuclear localization signals (NLS)" CRITICAL REVIEWS IN EUKARYOTIC GENE EXPRESSION, vol. 3, no. 3, 1993, pages 193-227, XP002113642 page 195 - page 201; table I page 205, left-hand column, paragraph 2 - page 208, right-hand column, paragraph 2 -----	34-40, 42-50, 52-63

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

06/18/2003, EAST Version: 1.03.0002

INTERNATIONAL SEARCH REPORT

In.. national application No.
PCT/US 99/06675

B x I Observati ns where certain claim were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 18, 21, 24, 27, 29-31. as far as in vivo methods are concerned and 19, 22, 23, 25, 28 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [X] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

06/18/2003, EAST Version: 1.03.0002

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-3 all totally; 18-33 all partially

   Nucleic acid having the nucleotide sequence as in Seq.ID:2, encoding the polypeptide having the aminoacid sequence as in Seq.ID:36. Polypeptide having the aminoacid sequence as in Seq.ID:36. Application of said nucleic acid, or polypeptide in therapy and diagnostics.

2. Claims: 4 totally; 18-33 all partially

   As invention 1 but concerning Seq.ID:3.

3. Claims: 5-17 all totally; 18-33, 35-41, 43-65 all partially

   Nucleic acid having the nucleotide sequence as in Seq.ID:4, or fragments thereof, as comprised in Seq.ID:1, 5, 6, 7, 8, 9, 19. Corresponding encoded DRM polypeptide, or fragments thereof, as comprised in Seq.ID:29, 30, 31, 32, 33, 34, 35. Application of said nucleic acid, or polypeptide in therapy and diagnostics.

4. Claims: 34, 42 all totally; 35-41, 43-65 all partially

   A fusion polypeptide comprising a DRM protein, or fragments thereof, and a Green Fluorescent Protein, having the aminoacid sequence as in Seq.ID:29, 30, 31, 32, 33, 34, 35, and method of production thereof. Nucleic acid encoding said fusion proteins having the nucleic acid sequence as in Seq.ID:1, 5, 6, 7, 8, 9, 19.

INTERNATIONAL SEARCH REPORT

information on patent family members

Inter nal Application No
PCT/US 99/06675

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9726333 A	24-07-1997	US 5874304 A	23-02-1999
		AU 1750297 A	11-08-1997
		CA 2243088 A	24-07-1997
		EP 0874903 A	04-11-1998
WO 9837195 A	27-08-1998	AU 6177998 A	09-09-1998
WO 9833918 A	06-08-1998	AU 6267098 A	25-08-1998

06/18/2003, EAST Version: 1.03.0002